

Effects of Text Length and Question Type on Test-takers' Performance on Fill-in-the-blank Items in Korean CSAT

Minryoung Bae*

(Seoul National University)

Byungmin Lee

(Seoul National University)

Bae, Minryoung, & Lee, Byungmin. (2018). Effects of text length and question type on test-takers' performance on fill-in-the-blank items in Korean CSAT. *English Teaching*, 73(4), 149-174.

This study examines the effects of text length and question type on Korean EFL readers' reading comprehension of the fill-in-the-blank items in Korean CSAT. A total of 100 Korean EFL college students participated in the study. After divided into three different proficiency groups, the participants took a reading comprehension test which consisted of 4 reading passages (2 short and 2 long) from the Korean CSAT, followed by multiple-choice fill-in-the-blank questions and open-ended inference questions. The longer version of the passages was made from its originally restored version in which one or two paragraphs were added. The results showed that the college students performed better on the long passages than the short ones. In addition, the college students' reading comprehension test performance was affected differently depending on the type of questions. The findings of the study provided implications on how to select and construct reading passages for high-stake nationwide examinations, such as the Korean CSAT.

Key words: reading comprehension, text length, question type, inference, fill-in-the-blank item, Korean CSAT

1. INTRODUCTION

Assessing reading ability is an important part of a large standardized English test. Especially, the reading section in high-stake tests, such as the College Scholastic Ability

*Minryoung Bae: First author; Byungmin Lee: Corresponding author

Test (CSAT) plays a critical role in determining test-takers' overall English proficiency scores. Therefore, while it is important to use types of questions that can differentiate students' English abilities by means of these tests, it is also essential to consider the validity of these tests in accordance with the original purpose of the test. In other words, the construct of the reading comprehension ability should be clearly identified and validly assessed in order to properly interpret the students' performance in the standardized reading comprehension test (Bachman, 1990; Bachman & Palmer, 2010; Grabe & Jiang, 2014; Hsu, 2008; Salehi, 2011; Zhang, 2016).

Even though the CSAT has many functions, the goal of discriminating the students' reading ability according to their scores seems to be dominant in developing the reading section. To ensure the reliability and credibility of the test, it depends greatly on a multiple-choice type of questions (Alderson, 2000; Kim & Chon, 2014). In addition, the types of questions in the CSAT are organized into a fixed pattern by repeatedly using the similar types of questions. Another characteristic is that the passages in the test are somewhat short in length. Even though they are usually extracted from a variety of authentic sources, such as a journal article, technical research report or professional literature, the passages may range approximately from 100 to 200 words. For the CSAT, reading texts are usually more or less 148 words average (Kim, 2016).

However, this practice of reading assessment is not desirable considering a recent development and progress in reading theories and assessment. There have been various efforts to improve the reading evaluation reflecting the recent development of reading theories. There is a tendency to adopt longer and more authentic texts both by increasing amount of reading required and by reflecting reading tasks expected in the real-life context (Chapelle, Enright, & Jamieson, 2008; Ferrer, Vidal-Abarca, Serrano, & Gilabert, 2017; Grabe & Jiang, 2014; Keenan, Betjemann, & Olson, 2008; Liu, Schedl, Malloy, & Kong, 2009; Sainsbury, Harrison, & Watts, 2006). Second, it is emphasized to assess reading skills diversifying the question types as well as reading comprehension tests (Alderson, 2000; Kang, 2005; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Ozuru, Briner, Kurby, & McNamara, 2013; Ozuru, Rowe, O'Reilly, & McNamara, 2008).

Against this backdrop, the fill-in-the-blank type of questions, assumed to play the most critical and discriminating role in the English section of the CSAT test in Korea (Nam, 2015), is often perceived to be the most challenging questions by many students (Liu, 2009; Yamashita, 2003). The unique feature of the fill-in-the-blank type of questions as opposed to other types of questions is usually to extract one or two segments from an original text. For example, a question in 2014 CSAT was extracted from an article that runs almost 14 pages in total. Not only is the fill-in-the-blank question difficult to comprehend, but also it is somewhat out of context to understand the text. Also, such a kind of passage is relatively too short to understand not only some of the key expressions, but also an overall

message of the text. Furthermore, presenting such a passage based on the multiple-choice type of questions uniquely attracts test-takers' problem-solving strategies, such as applying logical problem-solving strategies and choice elimination strategies (Cohen, 1984; Connor & Read, 1978; Kern, 2000; Kim & Chon, 2014; Nam, 2015; Rupp, Ferne, & Choi, 2006; Wolf, 1993a), rather than reading the entire passage presented. Given that a variety of high-stakes decisions are made based on the test scores of this type of questions, there is a need for further research, investigating the validity of this type of questions.

Thus, the present study attempts to examine the multiple-choice fill-in-the-blank type of reading assessment in the CSAT in terms of its construct validity, and how it is used to assess students' English reading ability. In order to achieve this purpose, the amount of passage, specifically, passage length and questions types are manipulated to examine how they affect the Korean EFL readers' reading comprehension performance.

2. REVIEW OF THE LITERATURE

2.1. Effects of Text Length

Research has shown that a test-taker's performance is affected by a number of diverse factors including the test environment, methods of scoring, question types, topic familiarity, and others (Alderson, 2000; Bachman, 1990; Bachman & Palmer, 2010). Bachman (1990) discusses that test-takers' performance is affected by the characteristics of test methods. Among the test method facets, the text length is also a part of the facets of the test input. Since the difficulty of reading tests is determined by an interaction between test items and texts used in the reading comprehension tests, choosing the appropriate texts for assessing the learners' reading abilities should be considered as crucial (Ozuru et al., 2008).

Reading passages used in most standardized comprehension tests tend to be short. Due to the brevity of the short passages, the organizational structures of short passages are often less apparent than the reading texts in real-life settings (Commander & Stanwyck, 1997; Derrick, 1953; Valencia & Pearson, 1988). One outcome is that even though the short reading passages do not contain any problems in terms of linguistic features, it could fail to lead readers to build a coherent meaning of the text and therefore an appropriate situation model (Grabe, 2009; Kintsch, 1988; Sternberg, 1991).

As consensus has been reached among reading researchers, reading comprehension is the reader's active processing of constructing a coherent mental representation of the text by actively utilizing the text-based information such as word, sentence, discourse, and the reader-based information such as background information (Just & Carpenter, 1987; Kintsch, 1988; Perfetti, 1985; van Dijk & Kintsch, 1983). However, when it comes to the

short reading passages used in the context of standardized reading comprehension tests, these passages may not provide enough contexts and thus, could be limited in assessing normal reading performance. This lack of structure often makes readers spend more time to figure out the meaning of the text as a coherent whole, rather to depend on their background knowledge, not on the given passages (Valencia & Pearson, 1988). Therefore, they could be often limited to measuring a higher-level of reading comprehension processing, such as inference generation, and mental representation of the text (Kintsch, 1988; Kintsch & Rawson, 2005).

The claims in favor of using longer texts are in agreement with the authenticity argument in which students are often required to read lengthy texts for college-level academic coursework (Alderson, 2000; Flippo & Schumm, 2000; Valencia, Hiebert, & Afflerbach, 1994). In other words, the passages in the tests should be representative of the reading tasks that students are going to encounter in the real-life setting, such as college classrooms (Magliano, Millis, Ozuru, & McNamara, 2007). According to their assertions, short reading passages typically used in the standardized reading tests do not attempt to evaluate the actual reading demand where they are required to read a longer text. Similarly, Flippo and Schumm (2000) contend that few reading materials used in the standardized reading tests resembles the typical college reading assignment in length. They also emphasize that it is necessary to include sustained passages of text in order to design the authentic reading tests.

On the other hand, passages that are too long can be burdensome and may end up assessing endurance rather than reading ability. As Chastain (1988) noted, language teachers intuitively think that the longer the passage, the more difficult it will become because the longer passages contain a greater number of ideas. For the second language learners, especially, those with lower proficiency, the longer reading passages would increase cognitive burden when they are required to solve problems as quickly as possible in the context of standardized reading tests. This claim is in line with the Language Threshold Hypothesis (Alderson, 1984; Clarke, 1980), which argues that the readers have to reach a certain “threshold” level of second language proficiency in order for them to fully demonstrate their knowledge obtained from their first language. Thus, when the readers are confronted with the lengthy text in the testing environment, they would not be able to fully demonstrate their reading ability due to their lower-level of second language proficiency.

A number of studies have empirically investigated the effect of the passage length on reading comprehension. As outlined in Table 1, these have defined text length in a quite different manner. Some have forty-four words and others have more than two thousand words. Moreover, the range of the “short” and “long” passages in terms of the length is not consistent among these studies. For example, a “long” passage in one study falls into the

category of "short" passage of other studies, and vice versa.

TABLE 1
Previous Studies Involving the Effect of Text Lengths

Study	Text Length (words)		Measures	Results
	Short	Long		
Todd & Kessler (1971)	44	140, 256	Free recall	No Significant difference
Rothkopf & Billington (1983)	1056	1709, 2689	Free recall	Short > Long (Significant difference)
Commander & Stanwyck (1997)	260	620	Free recall, MCQ, Self-rating	Long > Short (Significant difference)
Beach (2008)	604	2079	Cloze test	No Significant difference
Cha (1995)	100, 106	412, 423	MCQ	Long > Short (Significant difference)
Lee (1999)	100, 120	230, 270	MCQ	No Significant difference

Careful examination of these studies yields a few interesting points. Firstly, the results using the range of 100 and 300 words did not affect the readers' performance (Hashemi & Bagheri, 2014; Jalilehvand, 2012; Lee, 1999; Todd & Kessler, 1971). In other words, the length of the longer passage should be more than twice that of the short passage in order for readers to consider the two types of texts differentially. This speculation is in line with the previous research studies, which examined the length effect (Lee, 1999; Yi, 2013). Yi (2013) reviewed the literature concerning the text length in reading tests and noted several relevant issues.

Secondly, as Yi (2013) noted, the results of previously discussed studies have shown disagreements about the effects of longer passages on the reader's comprehension process. Although the length of the text in reading comprehension tests has been examined in a few previous studies in EFL contexts (e.g., Cha, 1995; Choi, 2011; Hashemi & Bagheri, 2014; Jalilehvand, 2012; Lee, 1999; Yi, 2013), they have shown mixed results concerning the appropriate length of reading passages. This fact provides support for the necessity for further studies in this area in order to fill the gap regarding the effect of the text length on reading comprehension tests.

Thirdly, unlike the studies conducted in English as a first language (L1) and English as a second language (L2) contexts, most of the studies conducted in EFL contexts listed above relied only on multiple-choice questions as a measure of reading comprehension. Specifically, the results demonstrate a pattern of insignificant differences except for a study conducted by Cha (1995). To the contrary, the findings in the studies using open-ended question formats (e.g., free recall) and multiple measures of reading comprehension appear

to produce significant differences (Alderson, 2000; Brantmeier, 2005; Commander & Stanwyck, 1997; Liu, 2009; Rothkopf & Billington, 1983).

2.2. Effects of Question Types

The second factor that has been considered to interact with the feature of the passage length in reading comprehension assessment is how the reader's comprehension is measured. It is important to identify the way in which the reader engages according to different question types (Kulesz, Francis, Barnes, & Fletcher, 2016; Ozuru et al., 2008; Shohamy, 1984; Wolf, 1993a, 1993b). A number of studies have investigated whether question type affects students' reading comprehension performance (Bridgeman & Rock, 1993; Campbell, 1999; Ozuruet al., 2007; Ozuru et al., 2008; Ozuru et al., 2013; Shohamy, 1984; Wolf, 1993a, 1993b). In these studies, multiple-choice and open-ended questions are the most popular question types that have been examined to assess test-takers' reading comprehension performance. Even though consensus has been reached among researchers that these two types of questions measure different reading processes (Brantmeier, 2005; Campbell, 1999; Carrell, 1991; Shohamy, 1984; Wolf, 1993a), it is still unclear how these processes are distinguished from each other (Magliano et al., 2007; Ozuruet al., 2013).

Most reading comprehension tests administered in the context of standardized tests consist of a series of short passages, followed by multiple-choice questions (Alderson, 2000; Garcia, 1991; Johnston & Afflerbach, 1982). In fact, several researchers have raised considerable doubts over the validity of multiple-choice questions as a measure of reading comprehension (Anderson, Bachman, Perkins, & Cohen, 1991; Cohen, 1984; Connor & Read, 1978; Hsu, 2008; Kern, 2000). They have criticized and raised the question of the construct validity of multiple-choice questions and its limited or sometimes invalid assessment of the actual reading processes that the readers should go through in reading. The criticism of the multiple-choice questions has been constantly raised by many reading researchers and assessment developers since 1980s (Cohen, 1984; Farr, Pritchard, & Smitten, 1990; Kern, 2000; Nevo, 1989; Pressley & Ghatala, 1988). Several recent studies to investigate the construct validity of multiple-choice questions (Damankesh & Babaii, 2015; Doe & Fox, 2011; Kim & Chon, 2014; Rupp et al., 2006; Salehi, 2011) revealed that the readers showed quite different behaviors compared to normal text reading. In response to this persistent criticism, test developers began adopting other alternative question types (e.g., cloze, matching, open-ended, recall, and summary, etc.) in a large-scale reading comprehension assessment (Alderson, 2000; OECD, 2003; Rodriguez, 2003).

According to a view of cognitive processing in reading, comprehension questions differ in the degree of processing demands imposed on readers; questions inducing text-based information might heavily depend on lower-level processing whereas inference-generation

questions may depend on higher-level processing. This speculation is based on Kintsch's (1988, 1998) Construction-Integration Model. Kintsch (1988, 1998) claims that there are two levels of comprehension (i.e., text-base and situation models). In order for readers to construct a coherent situational model of text, they need to use inferences to connect new propositions to networks of already active propositions and maintain a coherent network of ideas. Therefore, inferencing in reading comprehension requires more cognitive processing to build a consistent meaning of the text by linking the textual information presented in each sentence using prior knowledge. van Dijk and Kintsch (1983) argue that when the reader forms a coherent situation model, the inferencing process plays a key role in the integration of textual information. They distinguish between two types of inference: bridging inference and elaborative inference (van Dijk & Kintsch, 1983). Bridging inference is required to construct the coherence of the text. While readers read the text, they combine the concepts presented in each sentence to link them together. On the other hand, elaborative inference is based on prior knowledge of the readers in relation to both the concepts and the events described in the text (Magliano et al., 2007). Therefore, for bridging inferences, background knowledge other than the textual information presented in the text is not necessarily required, but for elaboration inferences, the reader should extensively draw background knowledge related to the topic of the text (Daneman & Hannon, 2001; Kulesz et al., 2016).

The results of readers' reading comprehension performance in a reading assessment, therefore, have not only the quantitative differences (i.e., what is the total score), but also the qualitative differences (i.e., what types of information processing or what type of cognitive or reading skills could be tested). For instance, when two readers receive the same total score in a standardized reading comprehension test, their reading abilities may differ qualitatively if one only got the literal items (i.e., text-base) right, and the other got inferential items (i.e., situation model) as well as the literal items right.

Consequently, in the present study, we wanted to explore the effect of both text length and question type in reading assessment, examining test-takers' reading behaviors. Therefore, we examine the effects of text length as a function of the types of questions on the reader's comprehension performance; that is, how the Korean EFL students' reading comprehension performance across the shorter and longer version of the text differs depending on both multiple-choice fill-in-the-blank questions and open-ended inference questions, which ultimately promote deeper cognitive processing.

Thus, the review led to the following research questions for this study:

1. To what extent does text length affect students' reading comprehension performance measured by the two different question types (multiple-choice fill-in-the-blank question and open-ended inference questions)?

2. How are three different proficiency groups' reading comprehension performances differentiated in two different versions of the passages (short vs. long) as measured by the two types of questions (MCQ vs. OEQ)?

3. METHODOLOGY

3.1. Participants

The participants are Korean college students who were enrolled for a co-ed university in Korea. We chose college students rather than high school students because the fill-in-the-blank items in the Korean CSAT are relatively difficult and challenging for most of the high school students. Due to the inherited difficulty in the fill-in-the-blank types of questions, these questions are often considered as the items for the discriminating the top-level students. So presumably what we can expect is that these questions could be not only inappropriate but also limited in revealing ordinary high school students' reading comprehension ability. After all, to examine the top-level high school students' reading behaviors in this test, relatively similar or a higher level of college students were recruited to verify that this item indeed differentiates the top-level students' English reading comprehension capability.

The college students were recruited from online bulletin boards of a university located in Seoul, Korea. The majors of the participants vary ranging from social to natural sciences: business administration, chemistry, computer sciences, medicine, architecture, nursing, and public administration. Prior to the main study, a preliminary screening¹ was conducted to control the topic familiarity of the passages used in this study. Thus, a total of one hundred Korean EFL college students participated in the study.

Table 2 describes the distribution of the participants. Some experimental studies have recognized the importance of properly grouping L2 participants at the time of the experiment to better understand their behaviors depending on their L2 proficiency levels (Chae & Shin, 2015; Lee, 2014). In the present study, the 100 college students were divided into three groups according to their TEPS scores: the advanced group with 800 or above (34 students), the upper-intermediate group with 670 to 800 (33 students), and the lower-intermediate group with 670 to 485 (33 students), respectively. For those who don't have

¹ The screening process was conducted in order to identify any participants who were not qualified for the experiment: those who studied the passages of 2015 and 2016 Korean CSAT (College Scholastic Aptitude Test) English section, who took the Korean CSAT tests within two years, and the students majoring in English literature and English education because they are more likely to teach English to high school students so that they might be familiar with the passages.

TEPS scores, their TOEIC scores have been converted into corresponding TEPS scores. The conversion was processed, using the conversion table² presented by the TEPS Council.

TABLE 2
Demographics of the Participants

Proficiency Level	Range of TEPS Scores	Average TEPS Score	Gender		Total
			Male	Female	
Advanced	943 - 800	864.4	17	17	34
Upper-intermediate	792 - 673	747.7	16	17	33
Lower-intermediate	668 - 485	610.1	18	15	33
Total			50	50	100

3.2. Instruments

For the present study, four short passages were selected from the reading sections of the 2015 and 2016 Korean CSATs. All four passages were used for multiple-choice fill-in-the-blank questions. To examine the effect of text length, a longer version of the passages was constructed by adding one or two following or preceding paragraphs from the original texts. Thus, the longer versions of the passages were created to have two or three times of the shorter ones in length (i.e., 171, 160, 159, 172, and 132 words for shorter passages vs. 367, 371, 525, 443, and 308 words for longer passages, respectively). Although there are not absolute standards dividing short from long passages, for the sake of simplicity, we refer to the shorter passages from the Korean CSAT as short passages and the longer version of the passages as long passages.

Next, test items were constructed for each passage in the following two question types: multiple-choice fill-in-the-blank questions, and open-ended inference questions (generating bridging and elaborative inferences). For the multiple-choice fill-in-the-blank questions, the same questions from the Korean CSATs were used. The open-ended inference questions, bridging and elaborative inference questions, were prepared to examine the readers' comprehension of the coherent relationship between adjacent sentences as well as throughout the whole passage. Even though there was a difference in length between the two versions of the passage, all of the reading comprehension questions were based on the parts that are common to both versions of the passages (see Appendix A, B and C for two types of inference generation questions). As a result, the shorter- and longer-version of the test asked the same questions despite the difference in text length. Finally, summary writing questions were requested for the purpose of inducing the participants to read the

² The conversion table is provided by the TEPS Council at <http://www.teps.or.kr>.

text thoroughly, but the scores of summary writing were not counted in the final analysis.

3.3. Procedure

In order to avoid the practice effect and to counterbalance the cognitive load and the administrative time on each test, two forms of reading comprehension test, Form A and Form B, were developed. The two forms of tests and the question types are shown in Table 3. The participants were randomly assigned to either one of the two groups in each proficiency level, with one group taking Form A and the other group taking Form B. The two groups were created using stratified random selection, bringing comparable two samples within each proficiency group.

To eliminate the effect of time pressure, the participants were not assigned any time limit during the reading task. Although there was no time limit to complete the task, it took most students 40 minutes to complete the given task. Also, all the open-ended inference questions could be answered in Korean to get rid of any negative effects of using English in answering the questions (see Appendix C).

TABLE 3
The Two Formats of the Reading Comprehension Test

Passage	Text Length		Question Type			Score
	Form A	Form B	Multiple-Choice	Open-Ended		
				Bridging Inference	Elaborative Inference	
1	Short	Long	1	1	2	4
2	Long	Short	1	1	2	4
3	Short	Long	1	1	2	4
4	Long	Short	1	1	2	4
	Total score		4	4	8	16

3.4. Scoring

All the multiple-choice fill-in-the-blank questions were counted one point in each question and the open-ended inference questions were assigned either one or two points depending on question types, assigning one point for the bridging inference questions and two points for the elaborative inference ones. For the bridging inference items, one point was given for correct answers without any partial scoring. For the elaborative inference items, on the other hand, the points given to each answer ranged from 1 to 2. Two points were given for the answers which exactly match the model answers and one point for partial or near-right answers in order to give credit for the students' incomplete, but meaningful understanding of the passages. Thus, each participant theoretically gets a total

of 16 points as a perfect score from all the reading comprehension questions.

The students' response data for open-ended inference questions were scored based on the model answers. The answers were constructed by the first author. In order to examine the validity of the model answers, two additional native speakers were invited. One of them had M.A. in applied linguistics and the other had an M.A. in English literature. Both of them were doctoral students in English education at a university. Any disagreement between them was resolved through discussion. Finally, a model sample answer sheet for the open-ended inference questions was prepared.

To assess the reliability of marking open-ended inference questions, a Korean judge was invited. The Korean judge was a graduate student studying English education at a university and had five years of teaching experience at a Korean high school. The open-ended inference questions were scored by the first author, and then the invited rater scored 20 samples out of 100. The inter-rater reliability between the two judges in scoring the twenty samples of the open-ended inference questions was calculated. The inter-rater reliability between the two raters was significantly high ($r = .827, p < .05$).

4. RESULTS AND DISCUSSION

4.1. Effects of Text Length and Question Type

With regards to the first research question, a two-way repeated measures analysis of variance (ANOVA) was conducted. Dependent variables were test-takers' reading comprehension scores and independent variables were text length and question type. Table 4 provides mean and other descriptive scores for each group's performance on the short and the long passages in relation to the two question types. Figure 1 illustrates how the three different proficiency groups performed on the four different reading comprehension formats; each of the two different versions of the passages (short and long) was measured by the two types of questions (MCQ and OEQ). Table 5 gives a summary of the repeated measures two-way ANOVA for the effects of text length and question type.

A significant interaction effect between text length and question type was detected for the advanced group [$F(1,33) = 12.261, p = .001$] and for the upper-intermediate group [$F(1,32) = 11.435, p = .010$]. This shows that in both of the top-level proficiency groups, the effect of text length on the readers' performance was different depending on the type of questions they are given. For the lower-intermediate group, on the other hand, no significant interaction effect between text length and question type was detected [$F(1,32) = 1.578, p = .218$]. In this group, the main effect of text length was not detected, indicating that the lower-intermediate level students were not influenced by the text length. It means

TABLE 4
Descriptive Statistics for Students' Reading Performance

Proficiency Groups	N	Question Types	Text Length	Mean	SD	Minimum	Maximum
Advanced	34	MCQ	Short	1.47	.748	0	2
		MCQ	Long	1.35	.812	0	2
		OEQ	Short	3.55	1.078	2	6
		OEQ	Long	4.47	1.186	1	6
Upper-intermediate	33	MCQ	Short	1.15	.712	0	2
		MCQ	Long	1.12	.819	0	2
		OEQ	Short	2.96	1.185	1	5
		OEQ	Long	4.00	1.030	2	6
Lower-intermediate	33	MCQ	Short	0.93	.939	0	2
		MCQ	Long	1.09	1.090	0	2
		OEQ	Short	3.27	1.375	1	6
		OEQ	Long	3.81	1.333	1	6

Note. MCQ: Multiple-Choice Questions, OEQ: Open-Ended Questions

FIGURE 1
Results of the Students' Reading Comprehension Performance by Proficiency Group

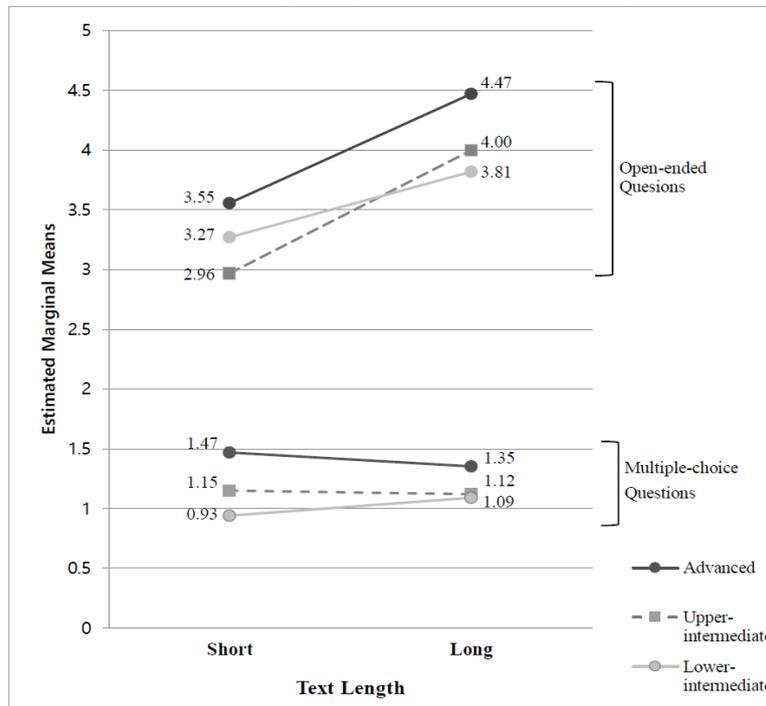


TABLE 5
Results of the Two-Way Repeated Measures ANOVA for the Effects of
Text Length and Question Types

Proficiency Group	Source	SS	df	MS	F	p	Partial η^2
Advanced	TL	5.360	1	5.360	4.548	.040*	.121
	QT	230.360	1	230.360	347.281	.000***	.913
	TL*QT	9.007	1	9.007	12.261	.001**	.271
	Error	24.243	33	.735			
Upper-intermediate	TL	8.250	1	8.250	8.250	.010**	.191
	QT	182.008	1	182.008	358.582	.000***	.918
	TL*QT	9.280	1	9.280	11.435	.002**	.271
	Error	24.970	32	.780			
Lower-intermediate	TL	4.008	1	4.008	3.109	.087	.089
	QT	211.280	1	211.280	147.074	.000***	.821
	TL*QT	1.280	1	1.280	1.578	.218	.047
	Error	25.970	32	.812			

Note. TL: Text Length, QT: Question Type, Significance level: * $p < .05$, ** $p < .01$, *** $p < .000$

that the lower-intermediate group performed similarly regardless of whether they were given either short or long passages.

In order to understand the significant interaction between text length and question type detected in the advanced and the upper-intermediate groups, we conducted tests for the simple effects of text length, which are shown in Table 6.

TABLE 6
Pairwise Comparisons for the Interaction between Text Length and Question Type
for the Advanced and the Upper-intermediate Groups

Proficiency Group	Question Type	Text Length		Subjects (F1)		p
		(A)	(B)	Mean Difference (B-A)	Std. Error	
Advanced	MCQ	Short	Long	-.118	.178	.513
	OEQ			.912	.284	.003**
Upper-intermediate	MCQ	Short	Long	-.030	.206	.884
	OEQ			1.030	.266	.001**

Note. MCQ: Multiple-Choice Questions, OEQ: Open-ended Questions, Significance level: ** $p < .01$

The results revealed that the main effect of text length was statistically significant only for the open-ended inference questions, not for the multiple-choice fill-in-the-blank questions. For the advanced group, the readers' comprehension performance measured by the open-ended inference questions was better in the long passages than the short ones [Mean Difference = .912, $p = .003$]. For the upper-intermediate group, the students' scores were higher in the open-ended inference questions when the long texts were given [Mean Difference = 1.030, $p = .001$]. In the multiple-choice fill-in-the-blank questions, on the

other hand, the difference of mean scores between the short and the long passages was not statistically significant. That is, for the advanced and upper-intermediate groups, text length positively influenced these groups' performance in the open-ended inference questions, not in the multiple-choice fill-in-the-blank items.

The results can be interpreted as follows: providing additional and sometimes redundant information had a positive effect on answering the open-ended inference questions for the students with the advanced and the upper-intermediate English proficiency, whereas giving additional information did not contribute much to answering the multiple-choice fill-in-the-blank questions. Thus, this finding implies that when the test-takers were asked to answer the multiple-choice fill-in-the-blank questions, which only requires a selective response, a longer text may not be necessarily helpful for the readers to find the most likely option among the five choices.

The findings also suggest that the actual reading comprehension process that test-takers go through while they find the answer in the multiple-choice fill-in-the-blank questions does not seem to match with the expectations held by test developers who would assume that the test-takers read the entire passages to answer the question. If the test-takers focused on the overall understanding of the given text, the additional information provided should have contributed to the improvement of the students' reading comprehension, but not answering the questions. As the results display, it cannot be assured that the additional information was useful to answer the multiple-choice fill-in-the-blank questions, and it would be difficult to assume that the readers might have attempted to read the additional information in the long passages to find an answer to the questions.

The results of the study are consistent with findings in Rupp et al. (2006), which showed that the examinees might utilize logical problem-solving or elimination strategies only to deduce the most appropriate option in the multiple-choice questions. In the same vein, the top-level college students could have solved the multiple-choice fill-in-the-blank questions without necessarily reading the entire text, as shown in their similar scores in both the short and the long passages. However, when the long passages were provided, their degree of understanding of the text improved, which was reflected by the better scores only in the open-ended inference questions.

Rupp et al. (2006) cast a doubt on the validity of multiple-choice questions in reading comprehension assessment, claiming that "a blanket statement such as 'MC questions assess reading comprehension' is nonsensical for any test" (Rupp et al., 2006, p. 470). In a recent study, Kim and Chon (2014) also cautioned the limited role of the multiple-choice questions as a valid measure of the readers' actual reading comprehension.

The results regarding the two different types of questions might indicate that the two question types, multiple-choice fill-in-the-blank and open-ended inference questions differ in the processes in which the test-takers engage while answering the questions (Gordon &

Hanauer, 1995; Ozuru et al., 2013; Wolf, 1993a). In other words, the text length of the reading comprehension tests does not seem to independently affect the test-takers' text comprehension, but the text feature might have a different effect on the readers' reading performance depending on the assessment tasks that they are engaged in.

The results also indicate that the test-takers set different reading goals to answer the reading comprehension questions (Doe & Fox, 2011; Rupp et al., 2006; Wolf, 1993a). The notion that the two question formats assess different reading processes can also be found in the previous studies, such as Ozuru, Brine, Kurby, and McNamara's (2013) research. They concluded that open-ended and multiple-choice format questions assess different aspects of reading comprehension. Therefore, there can be no single perfect evaluation method that measures the reader's reading ability (Alderson, 2000; Liu, 2009; Wolf, 1993b). When assessing the test-takers' reading comprehension capability, it might be appropriate to measure the reader's reading skills by means of various reading assessment methods.

Interestingly, the effect of text length as a function of the type of questions in this study was not detected in the lower-intermediate group. In both of the two reading comprehension measures, the lower-intermediate group performed similarly regardless of whether the given passages were short or long. One plausible explanation for these results could be due to the difficulty of the reading passages used in the experiment. It should be noted that all reading passages used in the study were taken from the actual 2015 and 2016 Korean CSAT for high school seniors. Given that the lower-intermediate group consists of college students who obtain the average TEPS score of 610 (within the range of 668 to 485), it seems logically possible to deduce that their reading behavior is likely to be similar or much higher compared to most of the high school seniors or graduates who actually take the test. Rather, providing additional information in the reading passages might have caused this level of readers to be distracted due to their limited English reading proficiency. It could also be interpreted that the reading passages used in this experiment might be too difficult for them to comprehend appropriately.

4.2. Effects of Text Length and Question Type for Proficiency Groups

To answer the second research question, how the three English proficiency groups were differentiated in each of the four different sets of reading comprehension tests, one-way analysis of variance (ANOVA) was conducted. Separate one-way analyses of variance (ANOVA) indicated a statistically significant group difference as shown in Table 7. Interestingly, when the short passages were provided for the multiple-choice fill-in-the-blank questions, there was a statistically significant difference among the three proficiency groups [$F(2, 97) = 4.600, p = .012$]. In addition, there was a barely detectable statistically significant difference in the open-ended inference questions when the long passages were

given [$F(2, 97) = 2.690, p = .073$]. However, no statistically significant differences were detected in the other two reading assessment formats.

TABLE 7
ANOVAs on Proficiency Group by Text Length and Question Type

Source		<i>df</i>	MS	<i>F</i>	<i>p</i>
MCQ-Short	Between groups	2	2.399	4.600	.012*
	Within groups	97	.522		
	Total	99			
MCQ-Long	Between groups	2	.691	1.082	.343
	Within groups	97	.639		
	Total	99			
OEQ-Short	Between groups	2	2.906	1.959	.147
	Within groups	97	1.483		
	Total	99			
OEQ-Long	Between groups	2	3.810	2.690	.073
	Within groups	97	1.416		
	Total	99			

Note. MCQ: Multiple-Choice Questions, OEQ: Open-ended Questions, Significance level: * $p < .05$

To find out where significant group differences exist, a post-hoc pairwise comparison using Tukey's HSD test was conducted. As shown in Table 8, in the reading test with the short passages measured by the multiple-choice fill-in-the-blank questions, there was a significant difference between the advanced and the lower-intermediate groups ($p = .009$). On the other hand, in the open-ended inference questions with the long passages, there was a marginally statistically significant difference between the advanced and the lower-intermediate groups ($p = .069$), which suggested that the advanced group outperformed the lower-intermediate group in the open-ended inference questions with the long passages.

TABLE 8
Post Hoc Test: Pairwise Comparisons by Group

Source	Group (A)	Group (B)	Mean Difference (A-B)	Std. Error	<i>p</i>
MCQ-Short	High	Mid	.319	.176	.172
		Low	.531	.176	.009**
	Mid	Low	.212	.177	.460
MCQ-Long	High	Mid	.231	.195	.464
		Low	.262	.195	.376
	Mid	Low	-.231	.195	.987
OEQ-Short	High	Mid	.589	.297	.123
		Low	.286	.297	.603
	Mid	Low	-.303	.299	.572
OEQ-Long	High	Mid	.470	.290	.243
		Low	.652	.290	.069
	Mid	Low	.181	.292	.809

Note. High: Advanced, Mid: Upper-intermediate, Low: Lower-intermediate, Significance level: ** $p < .01$

The reason for these findings might be interpreted that a type of questions may influence how learners with different English proficiency interact with the characteristics of the text. In other words, depending on the characteristics of the text given, the readers' understanding of the text might be differently affected by a certain type of questions. Thus, if we want to measure the student' overall understanding of the text, we should measure their reading comprehension through a type of questions which directly induce their understanding of the entire passage, not a type of questions which only tap into limited understanding of the passage.

Another possible interpretation of the results would be that the test-takers with the highly advanced English proficiency seem to be able to adjust their purpose of reading in the standardized reading comprehension test (Shohamy, 1984; Wolf, 1993a). In responding to multiple-choice questions, the advanced-level students might be able to utilize some metacognitive test-taking strategies, such as discriminating the best option from distractors, paying selective attention to the portion of the passage which is related to find the answer, or rereading the critical part (Kim & Chon, 2014; Rupp et al., 2006). Also, they could utilize the options followed by the multiple-choice questions, which, in turn, provide richer information to find an answer (Gorden & Hanauer, 1995). Finally, the multiple-choice fill-in-the-blank questions presented in this experiment are similar to TEPS in many respects. Therefore, the highest group that received the best scores in TEPS is likely to have acquired useful testing strategies for this type of multiple-choice questions. In other words, the advanced group might have utilized contributory strategies in order to accomplish the purpose of the testing (Kim & Chon, 2014; Zhang, 2016). Thus, it is possible to say that the multiple-choice fill-in-the blank items using the short passages distinguished these three groups of subjects as opposed to the other three testing formats.

Similarly, when answering the open-ended inference question, students need to read and understand the whole text thoroughly. For the highly proficiency readers, presumably reaching TEPS 800 or above, the difficulty dealing with the open-ended inference questions might have been alleviated by the additional text given to them. For the lowest group, within the range of TEPS 600 or lower, however, the long passages with additional information might have caused them some additional burden.

5. CONCLUSION

The present study was designed to examine the validity of the fill-in-the-blank items of the Korean CSATs in terms of how they reflect test-takers' reading comprehension ability. The results point to the relative benefits of the lengthier text of reading comprehension tests only in the open-ended inference question, but not in the multiple-

choice fill-in-the-blank items. The positive effects of giving more textual information were reflected in the reading comprehension performance, measured by only the open-ended inference questions for the two top-level proficiency groups. The findings of the study indicate that the readers with a relatively higher L2 proficiency could better understand the overall discourse of the long passages, which, in turn, improved their reading comprehension performance measured by the open-ended inference type of questions.

Despite some meaningful and significant findings, there are a few caveats which should be mentioned here. One is that this study investigated the performance of the college students showing high performance not only in English but also in academic performance. Thus, the findings drawn from such a population may not be directly applicable to other groups of Korean students. For example, to get more direct results with regards to the effects of the two main factors and their interaction on the top-level high school students, it is suggested that future studies recruit this group of high school students to measure their reading performance with the same format.

Another caveat of the study lies in not controlling the time students spent reading the text. Clearly, having students answer questions without time limit might make it difficult to generalize the findings of the study in a real testing situation. However, the focus of the present study was to examine the effects of text length on the students' performance on the fill-in-the-blank items depending on the types of questions. In this respect, if the experiment had been administered with time limit, different results would probably have been obtained. However, considering the fact that their English proficiency levels in this study were much higher than the average high school students, the correct answer rate was still relatively low. Therefore, it is important to consider whether the level of difficulty of the passages is appropriate for ordinary Korean high school students. Also, it should be pointed out whether the reading passages contain enough information to appropriately understand them.

The results of the present study lead to the following pedagogical implications. In the standardized reading comprehension tests administered in Korea, a series of short reading passages are dominantly exploited. These short reading passages often do not have natural discourse structures reflected in a longer passage or academic textbooks that students read in colleges and universities. Even if the shorter passages are used to develop to expand students' vocabulary knowledge and to improve their syntactic parsing skills, it often makes it difficult for readers to construct a coherent representation of the text. Consequently, readers often struggle to understand the meaning of the unnaturally short and out-of-context texts and they may simply give up on reading the passage and may end up taking a wild guess based on their background knowledge which is not quite relevant to the text.

The fact that the multiple-choice fill-in-the-blank questions discriminated the students' proficiency level better than the other testing formats is a remaining issue that should be examined further, using various research methods. It should be noted that the present study adopted a quantitative approach to investigate the influence of the two factors on the students' reading performance. Thus, the students' reading comprehension behaviors are based on their testing results. Further studies might probe this issue by utilizing more diverse research methods, such as think-aloud, stimulated recall interview, or eye-tracking to gain a better understanding of the test-takers' actual reading and problem-solving behaviors in this type of questions.

The findings of the present study clearly suggest that it is essential to know how the amount of text influences reading comprehension in a reading comprehension assessment and how differently questions types reflect the outcome of reading comprehension. In order to interpret the performance of assessment as an indicator of the test-taker's reading abilities, the test developers should first clearly identify and define the construct of reading comprehension abilities. If the test developers' expectations do not match with the actual processes that test-takers undergo during testing, it is difficult to make any meaningful inferences based the test scores. Especially, if the results of the reading scores have a serious consequence on the test-takers like high school graduates, the reading comprehension section of the Korean CSAT, for example, need to incorporate more valid reading comprehension questions, especially adopting a longer passage rather than a truncated short one with information incoherently compacted in the passage.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading*. New York, NY: Cambridge University Press.
- Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1-27). London: Longman.
- Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, 8(1), 41-66.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- 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, Tempe, AZ.

- Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal*, 89(1), 37-53.
- Bridgeman, B., & Rock, D. A. (1993). Relationships among multiple-choice and open-ended analytical questions. *Journal of Educational Measurement*, 30, 313-329.
- Campbell, J. R. (1999). *Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension*. Unpublished doctoral dissertation, Temple University, Philadelphia, PA, USA.
- Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, 12(2), 159-179.
- Cha, K. A. (1995). The effect of text length in the diagnosis of reading comprehension. *Journal of the Applied Linguistics Association of Korea*, 8(1), 249-275.
- Chae, E.-Y., & Shin, J.-A. (2015). A study of a timed cloze test for evaluating L2 proficiency. *English Teaching*, 70(3), 117-135.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Test score interpretation and use. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 145-186). New York, NY: Routledge.
- Chastain, K. (1988). *Developing second language skills: Theory and practice*. Orlando, Florida: Harcourt Brace Jovanovich.
- Choi, M. S. (2011). *Analysis of the factors affecting passage dependency of multiple-choice English reading comprehension tests*. Unpublished master's thesis, Ewha Womans University, Seoul, Korea.
- Clarke, M. A. (1980). The short circuit hypothesis of ESL reading or when language competence interferes with reading performance. *Modern Language Journal*, 64, 203-209.
- Cohen, A. D. (1984). On taking language tests. *Language Testing*, 1(1), 70-81.
- 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.
- Connor, U., & Read, C. (1978). Passage dependency in ESL reading comprehension tests. *Language Learning*, 28(1), 149-157.
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62-69.
- 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*, 130(2), 208-223.

- Derrick, C. (1953). Three aspects of reading comprehension as measured by tests of different lengths. *Research Bulletin* 53-8. Princeton, NJ: Educational Testing Service.
- Doe, C., & Fox, J. (2011). Exploring the testing process: Three test takers' observed and reported strategy use over time and testing contexts. *Canadian Modern Language Review*, 67(1), 29-54.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27(3), 209-226.
- Ferrer, A., Vidal-Abarca, E., Serrano, M. Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology*, 51, 404-415.
- Garcia, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, 26(4), 371-392.
- Gordon, C. M., & Hanauer, D. (1995). The interaction between task and meaning construction in EFL reading comprehension tests. *TESOL Quarterly*, 29(2), 299-324.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Grabe, W., & Jiang, X. (2014). Assessing reading. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp.185-200). Oxford, UK: John Wiley & Sons.
- Flippo, R. F., & Schumm, J. S. (2000). Reading tests. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 403-472). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hashemi, S. Z., & Bagheri, M. S. (2014). External factors influences on EFL learners' reading comprehension test performance. *International Journal of Language Learning and Applied Linguistics World*, 7(1), 150-166.
- Hsu, W. T. (2008). *EFL students' English language knowledge, strategy use and multiple-choice reading test performance: A structural equation modeling approach*. Unpublished Doctoral dissertation, University of Southampton, United Kingdom.
- Jalilehvand, M. (2012). The effects of text length and picture on reading comprehension of Iranian EFL students. *Asian Social Science*, 8(3), 329.
- Johnston, P., & Afflerbach, P. (1982). *The centrality of reading comprehension test questions and their validity*. Paper presented at the annual meeting of the New York State Reading Association, Kiamesha Lake, NY.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Kang, D.-M. (2005). A study on short-answer-question format in testing EFL learners' L2

- reading ability. *Language Research*, 41(4), 963-983.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300.
- Kern, R. G. (2000). *Literacy and language teaching*. Oxford: Oxford University Press.
- Kim, S. H., & Chon, Y. V. (2014). Test-taking strategies of L2 adolescent learners: Three multiple-choice items and L2 proficiency. *English Teaching*, 69(1), 61-90.
- Kim, S. J. (2016). *Analyzing the cohesion of the reading passages on standardized English tests: CSAT, TEPS, EIKEN and TOEFL*. Unpublished master's thesis, Seoul National University, Seoul, Korea.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209-226). Malden, MA: Blackwell Publishing.
- Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*, 108(8), 1078-1097.
- Lee, J.-K. (2014). Self-assessment of second language proficiency as a research tool. *English Language and Linguistics*, 20(2), 125-137.
- Lee, K. S. (1999). The effects of the number of questions per passage, the length of passage, and the topic familiarity on multiple-choice English listening and reading comprehension tests. *English Teaching*, 54(4), 327-351.
- Liu, F. (2009). The effect of three test methods on reading comprehension: An experiment. *Asian Social Science*, 5(6), 147-153.
- Liu, O. L., Schedl, M., Malloy, J., & Kong, N. (2009). *Does content knowledge affect TOEFL iBT™ reading performance? A confirmatory approach to differential item functioning*. Princeton, NJ: English Testing System.
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107-136). New York: Lawrence Erlbaum Associates.
- Nam, Y. K. (2015). Korean EFL students' strategy use in fill-in-the-blank inference items. *English Teaching*, 70, 81-107.
- Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension.

- Language Testing*, 6(2), 199-215.
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67(3), 215-227.
- Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40(4), 1001-1015.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Pressley, M., & Ghatala, E. S. (1988). Delusions about performance on multiple-choice comprehension tests. *Reading Research Quarterly*, 23, 454-464.
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40(2), 163-184.
- 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 Verbal Learning and Verbal Behavior*, 22(6), 667-681.
- Rupp, A. 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.
- Sainsbury, M., Harrison, C., & Watts, A. (2006). *Assessing reading: From theories to classrooms*. Slough: National Foundation for Educational Research.
- Salehi, M. (2011). Test taking strategies: Implications for test validation. *Journal of Language Teaching and Research*, 2(4), 850-858.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147-170.
- Sternberg, R. J. (1991). Are we reading too much into reading comprehension tests? *Journal of Reading*, 34(7), 540-545.
- Todd, W. B., & Kessler, C. C. (1971). Influence of response mode, sex, reading ability, and level of difficulty on four measures of recall of meaningful written material. *Journal of Educational Psychology*, 62(3), 229-234.
- Valencia, S. Hiebert, E. H., & Afflerbach, P. (1994). Definitions and perspectives. In S. Valencia, E. H. Hiebert, & P. Afflerback (Eds.), *Authentic reading assessment: Practices and possibilities* (pp. 6-21). Newark, DE: International Reading

Association.

- Valencia, S. W., & Pearson, P. D. (1988). Principles for classroom comprehension assessment. *Remedial and Special Education, 9*(1), 26-35.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wolf, D. F. (1993a). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal, 77*, 473-489.
- Wolf, D. F. (1993b). Issues in reading comprehension assessment: Implications for the development of research instruments and classroom tests. *Foreign Language Annals, 26*(3), 322-331.
- Yamashita, J. (2003). Processes of taking a gap-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.
- Zhang, L. (2016). Chinese college test takers' individual differences and reading test performance: A structural equation modeling approach. *Perceptual and Motor Skills, 122*(3), 725-741.

APPENDIX A

Example of Short Text

The concept of humans doing multiple things at a time has been studied by psychologists since the 1920s, but the term “multitasking” didn’t exist until the 1960s. It was used to describe computers, not people. Back then, ten megahertz was so fast that a new word was needed to describe a computer’s ability to quickly perform many tasks. In retrospect, they probably made a poor choice, for the expression “multitasking” is inherently deceptive. Multitasking is about multiple tasks alternately sharing one resource (the CPU), but in time the context was flipped and it became interpreted to mean multiple tasks being done simultaneously by one resource (a person). It was a clever turn of phrase that’s misleading, for even computers can process only one piece of code at a time. When ㉔ they “multitask,” they switch back and forth, alternating their attention until both tasks are done. The speed with which computers tackle multiple tasks that everything happens at the same time, so ㉕ comparing computers to humans can be confusing.

APPENDIX B

Example of Long Text

Multitasking is a lie. It's a lie because nearly everyone accepts it as an effective thing to do. It's become so mainstream that people actually think it's something they should do, and do as often as possible. We not only hear talk about doing it, we even hear talk about getting better at it. More than six million webpages offer answers on how to do it, and career websites list "multitasking" as a skill for employers to target and for prospective hires to list as a strength. Some have gone so far as to be proud of their supposed skill and have adopted it as a way of life. But it's actually a "way of lie," for the truth is multitasking is neither efficient nor effective. In the world of results, it will fail you every time. When you try to do two things at once, you either can't or won't do either well. If you think multitasking is an effective way to get more done, you've got it backward. It's an effective way to get less done. As Steve Uzzellsaid, "Multitasking is merely the opportunity to screw up more than one thing at a time."

The concept of humans doing multiple things at a time has been studied by psychologists since the 1920s, but the term "multitasking" didn't exist until the 1960s. It was used to describe computers, not people. Back then, ten megahertz was so fast that a new word was needed to describe a computer's ability to quickly perform many tasks. In retrospect, they probably made a poor choice, for the expression "multitasking" is inherently deceptive. Multitasking is about multiple tasks alternately sharing one resource (the CPU), but in time the context was flipped and it became interpreted to mean multiple tasks being done simultaneously by one resource (a person). It was a clever turn of phrase that's misleading, for even computers can process only one piece of code at a time. When ^(a) they "multitask," they switch back and forth, alternating their attention until both tasks are done. The speed with which computers tackle multiple tasks that everything happens at the same time, so ^(b) comparing computers to humans can be confusing.

APPENDIX C

Example of a Fill-in-the-blank Multiple-choice Question, an Open-ended Question, and Summary (translated into English)

1. Choose the option that best completes the passage.
 - ① expels the myth
 - ② feeds the illusion
 - ③ conceals the fact
 - ④ proves the hypothesis
 - ⑤ blurs the conviction
2. What does underlined ^(a) they refer to?
3. Why does the author insist that ^(b) "comparing computers to humans can be confusing"?
4. Write the summary of the passage in about 20 words.

Applicable levels: Secondary, tertiary

Minryoung Bae
Graduate Student
Department of English Language Education
College of Education, Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul 08826, Korea
Email: goodnews15@snu.ac.kr

Byungmin Lee
Professor
Department of English Language Education
College of Education, Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul 08826, Korea
Email: bmlee@snu.ac.kr

Received on September 29, 2018

Reviewed on November 17, 2018

Revised version received on December 12, 2018