

How does Text Availability Affect Performance in EFL Reading Comprehension Tests?

Jeongsoon Joh
(Inha University)

Joh, Jeongsoon. (1998). How does text availability affect performance in EFL reading comprehension tests? *English Teaching*, 53(3), 199-216.

The purpose of this study was to examine the effect of text availability while answering questions on performance in EFL reading comprehension tests, which has hardly been addressed in this area, in spite of its potential as an issue. The subjects were 648 Korean high school students. In order to examine whether and how text availability interacts with other variables, as well as its main effect, items are constructed both in subjects' native and target language, and four different test formats were used. Results of statistical analyses showed that students got significantly higher scores when they were allowed to refer to the test passages. However, the difference in mean scores resulting from text availability was not significant when students were tested in their native language, or when the task was very easy or too difficult. Memory constraints imposed by the text unavailable condition, combined with the difficulties resulting from limited target language proficiency, could explain these significant main and interaction effects. The findings in this study imply that we need to take into account the complexities involved in assessment of reading comprehension and in interpreting related research conducted in various contexts.

I. INTRODUCTION

Reading comprehension in a second or foreign language is a complicated cognitive activity that involves various factors, such as knowledge of syntax and lexicon of the language, interest and motivation, knowledge of the world, to name a few. Given this, variables related to assessment will surely add to the

complexity involved in performance on second/foreign language reading comprehension tests. Therefore, in educational settings, as well as in research context, we need to understand the factors that may affect performance in reading comprehension tests.

Reflecting this need, there have been some studies about the effects of language in which questions and/or answers are constructed, of types of test used to measure reading comprehension, and of text type (e.g., Kendall et al., 1980; Lee, 1986; Shohamy, 1984; Wolf, 1993). However, there have been few, if any, studies on the effect on reading comprehension performance, of text availability—that is, whether or not examinees are allowed to refer to the text upon which questions are based, while they are answering questions.

Text availability could be an issue of interest in that some researchers did not allow their subjects to refer back to the reading passages while they were working on the reading comprehension items, whereas most other researchers did. This inconsistency is probably based on differing underlying theories or models of reading comprehension. Researchers such as Johnston (1984) and Wolf (1993), for example, suggested that, for certain types of reading comprehension tasks, it would be desirable to have readers do without the text accompanying test items while answering questions.

Their suggestion, in one way or another, might be related to the common criticism on multiple-choice format as a measure of reading comprehension, in terms of validity. That is, it has been criticised in that it provides additional information about the text via stems and choices, (Bernhardt, 1991)¹⁾, and that test-takers can often rely on the recognition of a few key words in the passage, or matching them with those in stems and choices. As a result, unnecessary skills (or more commonly called test-taking strategies) such as matching surface structures of the text and test items, may intervene in the process, and therefore may discourage meaningful reading on the part of the readers, and decrease validity. It could be a way to amend this reported weakness of multiple-choice format if we do not allow examinees to refer to passages accompanying test items during test-taking, though not perfect or the only way to do so.

In fact, as mentioned earlier, some researchers (e.g., Kendall et al., 1980; Wolf, 1993) did not allow their subjects to refer to the reading passages once they began answering questions, but most others did. Findings from these

1) quoted in Wolf (1993)

reading research usually are accepted and applied without taking into account the difference in experimental conditions that might have affected the reported results. However, it is difficult to compare the findings from L2 reading comprehension research drawn from studies with different experimental conditions.

In addition, text availability conditions deserve attention because the difference in the cognitive demands of the two conditions (text available vs. text unavailable) is considerable, and this variable may supply relevant information on the question of whether information that was comprehended is retrievable (cf. Johnston, 1983).

In Korea, as well as in many other countries, almost all of the currently used reading comprehension tests, most of which use the multiple-choice format, provide examinees with the text upon which test items are based. Therefore the present study seeks to answer the questions of whether, and how much, text availability (i.e., whether the text is available for reference or not while answering questions) affects test scores of EFL learners, and whether text availability interacts with other variables that have been studied in the area of reading comprehension and/or its assessment, such as language in which test items constructed and task type used to measure reading comprehension.

II. METHOD

1. Research Questions

The research questions in this study were: 1) Do learners perform differently when they do not have stimulus texts to refer to while working on the reading comprehension tests?: 2) If so, does the effect interact with other variables associated with test features, such as the language in which questions and/or answers are constructed (language of assessment, hereafter), task type, and the degree of target language proficiency?

2. Participants

Participants were 648 Korean high school students (second graders) learning English as a foreign language. They were from eight high schools, two classes from each. There were 240 boys and 408 girls.

3. Design

A factorial design ($2 \times 2 \times 4$) was used for this study, which was fully crossed but not fully balanced, in that the size of each cell was not equal. For the variable of language, items are constructed both in Korean and English. Four types of test format were used: multiple-choice, open-ended, cloze, and summary task.

4. Materials

There were four reading passages, of about 500 words each, drawn from a reading text written by native speakers of English. According to the authors, their book was intended for intermediate level ESL learners or English native readers in developmental programs. The degree of difficulty (reading grade) of these passages was 9, calculated after the readability formula by McLaughlin (1969). According to McLaughlin, grades 13-16 indicate the need for college education, and grades 17-18, the need for graduate training.

For each passage, five multiple-choice items were constructed. One of them concerns main idea; another, inference; and the others, details. A total of 20 items, 5 per passage, were constructed for the multiple-choice format. Open-ended items were constructed simply removing choices from the multiple-choice items.

For cloze task, a summary of each passage was produced: first, three native speakers of English majoring in TESL summarized the four passages used in this study, and then the researcher compared those three different summaries, abstracted those idea units that were included in all three summaries, and created new summaries. These new summaries were examined by a native speaker of English to ensure their natural quality. By using a summary of the original text as the text for the cloze task, the cloze test became a post-reading task, similar and comparable to the other types of tasks being used in this study. A total of 20 blanks, 5 per passage, were left to be filled in. The filler for each blank was equivalent to the answer for each question in the multiple-choice and open-ended tests.

For summary task, students were instructed like the following: "Summarize what you read by using 120 to 150 words. Be sure to synthesize well what you read, deleting unnecessary details. Try to express in your own words rather than

to repeat phrases or sentences *per se* from the text. The focus should be on grasping the main idea of the entire text rather than listing details. You do not have to pay too much attention to the number of words suggested. It is just a guideline. What is important is to understand fully the main ideas and to express them in your own words if possible." The model summaries used for the cloze task were also used as a criterion against which to score the summary protocols.

Interview questions were constructed to obtain some data of qualitative nature. Questions included the kinds of difficulty students experienced while taking tests and strategies they used under a specific experimental condition.

5. Procedure

The main experiment was conducted in students' classrooms during their regular classes. In addition to the main experiment, a written interview was conducted outside their regular class hours in an effort to collect some qualitative data on what goes on in the students' mind while taking tests. The researcher administered all the procedures with the help of the instructors of the classes.

Students were randomly assigned to one of the 16 subgroups (i.e., two text availability conditions \times 2 languages of assessment \times 4 task types). The four passages were presented in a different order using Latin Square design, among participants, in order to avoid a possible order effect. Participants were given about five minutes of instructions on the nature of their tasks by the researcher before they began, such as what they were supposed to do with their tasks and how to complete them. The same written instructions were also given on the first page of their worksheet. Participants were given fifty minutes to read the four passages and answer the questions, including time for instructions. An additional ten minutes were allowed if they had not finished yet, but wanted more time.

Scores on all four tasks were scaled from 0 to 20 so that they could be compared across different tasks. Scoring of multiple-choice items was done simply by counting the number of correct answers, and therefore, scores were in the form of integers. For the open-ended and cloze tests, partial credits (of .5) were given, as well as giving full credits for correct answers, to those answers that have some potential as answers, but lacked something to be given full credits. As a result, score intervals for these task types were .5. In the case of the summary task, credits were given on the basis of idea units (or propositions)

that were the correct answers for each question in the other three tasks, in an effort to make the summary task equivalent to the other tests. First, each idea unit was assigned a certain percent so that the total could be 100. Then, penalty was assessed against distortion of the elements in the original text, intrusion of the elements not appearing in the original text, syntactic errors, and lack of natural flow as a discourse. A total score gained through these procedures was finally transformed so that a minimum score would be 0, and maximum, 20.

For the purpose of analysis, participants were classified into two different proficiency group after primary statistical analyses. The purpose was to see if there is any difference in the main effect or interaction effects among students of different degree of target language proficiency. English scores on a nationwide mock college entrance exam were used as the criterion of classification. Scores of the students in the lower proficiency group ranged from 20 to 71.3, and those in the higher group, from 74.4 to 100. The ANOVA (analysis of variance) and ANCOVA (analysis of covariance) was run first with the entire sample, and then, with each proficiency group separately.

After the main experiment, an interview was administered to 16²⁾ students drawn from the same sample, who represented eight different subgroups, i.e., two proficiency levels \times two languages of assessment \times two text availability conditions \times two students from each of these eight subgroups. Unlike the main experiment, the interview was administered outside the regular class hours. Interviewees took all four types of task used in this study, but only one passage was used for each task type. Finishing their tasks, they filled out interview questions as a sort of retrospective self-report.

III. RESULTS AND DISCUSSION

The results of descriptive statistics showed that the four tests used in this study had normal distribution of scores. An item analysis of the multiple-choice items showed that the percentage of students who correctly answered each item ranged from 22.5 to 95.6. Although students were randomly assigned to each

2) Although the researcher wished to interview more students, it was practically difficult because only one school allowed the researcher to meet students outside the regular classes.

subgroup, there was not a little difference in means on pretest scores among groups. In order to remove any pre-existing differences among groups that may influence the result of analysis, an ANCOVA was conducted, with pretest scores as a covariate. The observed and adjusted means are presented in Table 1, with standard deviations for the observed means.

TABLE 1
Observed/Adjusted Means and SD's for Text Availability × Language × Task

	Text Available		Text Unavailable	
	English	Korean	English	Korean
<u>Multiple-choice</u>				
Entire Sample	12.2 (4.8)	15.0 (3.3)	11.4 (3.8)	13.3 (3.6)
	13.0	15.8	11.5	13.7
Higher Level	15.8 (3.3)	17.4 (2.2)	13.8 (2.9)	15.2 (2.7)
Lower Level	9.6 (3.9)	13.9 (3.1)	8.8 (2.8)	11.2 (3.3)
<u>Open-ended</u>				
Entire Sample	9.8 (4.6)	12.1 (4.1)	7.1 (4.0)	11.0 (3.8)
	9.9	12.2	6.6	10.7
Higher Level	13.2 (2.9)	14.2 (2.9)	9.4 (2.5)	13.1 (2.8)
Lower Level	6.3 (3.2)	10.1 (4.3)	3.6 (3.3)	7.5 (2.1)
<u>Cloze</u>				
Entire Sample	11.1 (4.4)	12.7 (4.6)	8.7 (5.3)	12.0 (3.7)
	10.5	13.6	8.4	11.3
Higher Level	13.2 (2.9)	15.7 (2.6)	12.2 (3.7)	13.9 (1.9)
Lower Level	7.6 (4.4)	10.7 (4.9)	4.0 (3.2)	8.2 (3.5)
<u>Summary</u>				
Entire Sample	7.4 (3.7)	8.7 (4.7)	6.3 (4.5)	10.5 (3.5)
	6.8	9.0	6.6	9.8
Higher Level	9.8 (3.1)	12.1 (3.3)	9.0 (4.4)	12.2 (2.6)
Lower Level	4.9 (2.5)	5.6 (3.5)	3.7 (2.)	7.5 (2.6)
	11.2 (4.8)		10.2 (4.6)	
	n = 313		n = 335	
N = 648				

Note. Adjusted means are given in bold, and standard deviations are given in the parentheses.

Results of ANOVA revealed a significant main effect for text availability, and a significant two-way interaction effect for text availability by language of assessment. When the ANOVA was conducted only within each proficiency group, a significant main effect for text availability was found for both groups. However, there was no significant interaction effect within the higher proficiency group, while there was a significant interaction effect for text availability by task type within the lower proficiency group. The summarized results of the ANOVA are presented in Table 2.

TABLE 2
ANOVA Summary Table

Source of Variation	Sum of Squares	df	Mean Square	F	P
<i>Entire sample</i>					
Text availability	189.27	1	189.27	10.3	.001
Text availability x Language	74.02	1	74.02	4.2	.04
<i>Lower Proficiency Level</i>					
Text availability	231.73	1	231.73	19.2	.000
Text availability x Task type	118.47	3	39.49	3.3	.021
<i>Higher Proficiency Level</i>					
Text availability	196.02	1	196.02	23.6	.000

Note. Only significant effects are presented in this table.

The results of ANCOVA also showed a significant main effect for text availability. However, there was some difference in the type of significant interaction effect between these two analyses. In ANOVA, the interaction between text availability and task type was significant only within the lower proficiency group, but not with the entire sample. In ANCOVA, however, this interaction was significant with the entire sample put into analysis. When ANCOVA was conducted only with each proficiency level, the interaction between text availability and task type was significant with the lower

proficiency group, but not with the higher proficiency group, which is the same result as in ANOVA. In other words, the difference may be more apparent than real. Therefore, it seems that the source of significant interaction effect for text availability and task type came from the lower proficiency group. In addition, there was no significant interaction for text availability by language of assessment in ANCOVA, which was found significant in ANOVA. Judging from these results, it seems that the interaction between text availability and language of assessment was relatively weak, considering its significance level, .04, while the significant interaction for text availability by task type, on the other hand, seems quite a stable one.

1. Main Effect for Text Availability

The sample means and standard deviations for the main effect of text availability are shown in Table 3. As expected, students at both proficiency levels got significantly higher scores when they were allowed to refer back to the test passages while they were taking tests than when not. However, this main effect seems to need caution in interpretation because the variable of text availability revealed significant interaction effects with other variables, such as language of assessment or task type.

TABLE 3
The Means and SD's for Text Availability

		Text Availability	
		Text available	Text unavailable
Entire sample	Mean	11.2	10.2
	(SD)	(4.8)	(4.6)
Higherlevel	Mean	13.0	12.5
	(SD)	(3.5)	(3.5)
Lower level	Mean	9.0	6.9
	(SD)	(4.8)	(4.0)

Certainly, memory constraints under the text unavailable condition must have played a crucial role in the significantly different performance of students depending on the availability of the text while taking tests. Under the text

unavailable condition, they could not refer to the text looking for the information necessary to answer the questions correctly. Even though they may have understood the text when they read it the first time, they could not remember specific information or words to use when they had to answer the questions.

Although there was no statistically significant interaction between the text availability condition and the proficiency level of students, results indicated that students of lower proficiency would be more affected by the text availability than those of higher proficiency. That is, when text was not available for reference, test scores by students of lower proficiency slightly more deteriorated than that by students of higher proficiency. The reason that lower level students were more affected by the text unavailable condition might be explained in the following way. Students with limited proficiency of the target language generally have more difficulty in comprehending the text due to their limited knowledge of vocabulary and syntax of the target language, and therefore, may need more retrieval cues than higher level students. If they are even prohibited from referring to the text when they do not fully understand the text itself, they experience enormous difficulty because they are deprived of the opportunity to get some retrieval cues by referring to the text, at least for those parts they could comprehend but cannot remember.

In fact, interviewees reported this very problem, saying that they thought that they had understood well while reading the text, but once they had turned over the text, they could not remember the content, or they could not remember specific words even though they remembered the overall content, i.e., the macro structure of the text, using the term by van Dijk and Kintsch (1978). Because the students in this study who took the tests in the text unavailable condition knew, before they began, that they would not be allowed to refer to the text while taking the tests, they "read the text more intensively and tried to remember the content of the text" (as one interviewee claimed) than those who took the tests in the text available condition. However, because all students, regardless of the text availability condition, had to finish the tests within the same time limit, those who took the tests without the text available probably did not have enough time to rehearse the information in their working memory to the extent that it became a part of their long term memory. On the other hand, those students who took the tests with the text available surely could refer to the text whenever they wanted, looking for whatever information they did not remember clearly but found necessary to answer questions, or simply to

confirm the answers they had chosen based on their memory. Johnston (1983) postulated that comparing performance when text is available versus when it is unavailable during question answering would be a way to supply relevant information on the question of whether information which was comprehended is retrievable. The findings in the present study indicate that it is retrievable, but the extent of retrieval is affected by the availability of text during test-taking, a reader's proficiency in the language, and the nature of the test tasks at hand—the amount of retrieval cues provided by each test format. In addition, the text unavailable condition caused the readers to use different strategies to compensate for the difficulties resulting from greater demands and fewer retrieval cues inherent in the text unavailable condition, which will be discussed in a later section of this paper.

2. Interaction Between Text Availability and Language of Assessment

The results of statistical analysis revealed that the effect of text availability was different depending on the language in which questions were constructed and answers should be selected or produced. For a closer examination of the nature of the significant interaction between language of assessment and text availability, a post hoc test was conducted. The results revealed that when students were tested in their native language, their performance was not affected significantly by text availability. In other words, though students received higher scores when they could refer to the test passages, the difference in scores was not statistically significant when they were tested in their native language. On the other hand, the difference in mean scores resulting from the text availability condition was noticeably larger when students took the tests in the target language. Table 4 illustrates this point.

TABLE 4
Cell Means for Language \times Text Availability

	Text Available		Text Unavailable
Korean	12.1	=	11.7
	V		V
English	10.3	>	8.4

Note. The symbol "=" indicates non-significant differences, and "<," ">," or "V" indicates the direction of significant differences.

The reason that the text unavailable condition was more critical when students took the tests in the target language might be related to the fact that the target language generally requires more time to decode and encode than the native language, and therefore requires more time and effort for readers to store information coded in the target language in their long term memory. If students cannot refer to the text for the information necessary to answer the questions and have to depend just on their memory, it will surely increase the difficulty of test taking. By referring to the text, they might be able to check if their memory was correct, or find the information they need but do not remember. However, under the text unavailable condition, they are deprived of this possibility, and this, in turn, may lead to much lower achievement than when they take the tests in their native language, because difficulties imposed by the text unavailable condition would add to the difficulties resulting from the target language itself.

On the other hand, if they take the tests in their native language, the questions themselves will have the facilitating effect of clearly reminding them in their own language of what they have read, thus providing more retrieval cues. In the tasks that require production, students are likely to feel fairly comfortable with their native language because they will be able to express their ideas much better than in the target language, now that vocabulary or syntax will not be an obstacle for them, so long as they have understood the text well. Even when they have not understood the text fully, it would certainly be much easier for them to produce in their own language whatever of the text they had understood.

When the same post hoc test was conducted with each proficiency group, the higher group showed the same pattern as the entire sample. Different from this, scores of the students in the lower group were consistently higher with significance when the text was available, whether they were tested in the native language or target language. Therefore, the source of insignificant difference in means depending on the text availability condition when students took the tests in the native language seemed to come more from higher proficiency students, for the difference in means were significant for lower level students.

Though not statistically tested, results seem to suggest that certain students may not be affected seriously even though they cannot refer to the test passage

when they take tests in the native language. Many of the higher proficiency group students in this study had considerable command of English so that they could understand most of the text, though not completely. Therefore, questions given in the native language and producing answers in the native language surely had a facilitating effect for them. As a result, their performance did not show much decrement when they took tests in the text unavailable condition. Conversely, in the case of low proficiency students, though they were helped by the same facilitating effect of the language of assessment, their limited understanding of the text—surely due to their lack of knowledge in vocabulary and syntax in English—affected negatively their performance in the text unavailable condition.

3. Interaction Between Text Availability and Task Type

Both in the ANOVA and the ANCOVA, the interaction between text availability and task type was significant apparently in the lower proficiency group. Once again, a post hoc test was conducted to examine the nature of the significant interaction, and the result is shown in Table 5, with the same symbols as used in Table 4.

TABLE 5
Cell Means for Task Type and Text Availability (Lower Level)

	Text Available		Text Unavailable
Multiple-choice	11.9	>	10.1
Cloze	9.7	>	5.9
Open-ended	8.2	>	5.6
Summary	5.3	=	5.4

The finding that there was a significant interaction between text availability and task type seems to support the claim by Brown and Campione (1980)³⁾ that: the presentation of appropriate retrieval environments leads to access of materials previously 'forgotten': different testing situations provide different

3) quoted in Johnston (1983)

retrieval environments, and therefore, assessments of the availability of knowledge varies as a function of retrieval support in the testing context.

Whether the text was available or not, each task would have provided a different amount of retrieval cues within either of those specific conditions. For example, the multiple-choice test provided most, and the summary test least, and the open-ended or cloze somewhere between the first two. Consequently, students' relative performance on each test might have reflected the amount of retrieval cues that each test provided, and in addition this effect of task type as a function of retrieval support would have interacted with the text availability condition.

As shown in Table 5, the lower level students did not get significantly higher scores on the summary task when the test passages were available for reference than not, while they did on other types of task.

When the analysis was conducted including the entire sample at once, the multiple-choice and the summary tasks were hardly affected by the text availability condition. In other words, for these two types of task, the difference in mean scores was not significant between the text available condition and the text unavailable condition. The multiple-choice test was the easiest task for students at both proficiency levels in this study, and it provided in its nature most retrieval cues among the four task types via stems and alternatives. Therefore, it could be that the text unavailable condition did not impede performance significantly.

Then, how could one explain the non-significant difference in means between the text available and the text unavailable condition in the case of the summary task, when the means on the other three tests were significantly different depending on the text availability condition for the lower level students? Theoretically, when the text is not available for reference as questions are answered, there are greater demands on long term memory, cueing has a stronger function, and retrieval and organization skills are more important (cf. Johnston, 1983). On the other hand, when the text is available, readers concentrate more on recall of approximate location in the text, knowledge that one should look for it, search strategies, and logical reasoning skills. According to Johnston (1983), the summary task in the text available versus text unavailable conditions represents the difference between the ability to summarize and the ability to store, retrieve, organize, and summarize information. In the light of the

function of cueing, the summary task has no retrieval cues (readers must depend solely on their memory) if readers are not allowed to refer to the text while summarizing, whereas other types of tasks (i.e., multiple-choice, open-ended, and cloze tests) provide some retrieval cues by way of stems and/or alternatives—in the case of multiple-choice and open-ended tests, or the summary of the passage—in the case of cloze test. Therefore, the finding in this study that, for the lower proficiency level students, performance on the summary test was not significantly impeded in the text unavailable condition when the other tasks were, seems to be different from what theory would predict. A possible explanation for this discrepancy would be that the summary test was too difficult for them, so that the task was not much facilitated even when they could refer to the text.

4. Analysis of Interview Responses

There was no notable difference in strategies interviewees used among different types of tests when the text was available while taking the tests. However, there was a general tendency to read the text more carefully and intensively when they worked on open-ended or cloze formats than when on multiple-choice items. This seems natural considering the fact that multiple-choice items may give tessees some extra clues by reminding them of the content of the text. Generally, higher level interviewees used more strategies, were more detailed in describing the strategies they used, and more systematic in using strategies depending on the types of test, though not dramatically different, which was rarely the case with the lower level interviewees.

Most of the interviewees who worked under the text-unavailable condition mentioned their anxiety or nervousness about having to remember what they had read, whether the content of the text or specific words that they thought were important in making answers. All interviewees, regardless of their proficiency level, thought that they could have performed much better if they had taken the tests with the text available while taking tests. The results of statistical analysis showed that students actually performed much better under text available condition for both high and low proficiency groups.

As for strategy use, students who worked under the text unavailable condition, due to the limited access to test passages, generally tried to memorize

by repeatedly reading only what was considered important, whether it was words or main ideas. However, higher level students were different from the lower level counterparts in that some of them even focused on rhetorical organization of the text to facilitate comprehension and memorization. As in the case of students who worked under the text available condition, these students did not use dramatically different strategies depending on the type of test. Interviewees reported that they memorized, or tried to memorize, only the main stream of a text. Some lower level interviewees said that they used background knowledge in answering multiple-choice questions. While higher level interviewees focused on key words or phrases (for open-ended questions), and focused on understanding overall context of a given text, differentiated topic sentences from subordinate ones, and then organized in their mind how to put them in a discourse (for summary test), lower level ones memorized sentences or simply wrote down what remained in their memory, or simply gave up when they felt it was too difficult.

The results of analysis of students' response to interview questions revealed that students used different strategies when the text was not available for reference than when it was available. However, this difference and elaboration on strategy use under the text unavailable condition did not help their performance dramatically. In addition, dependence on prior knowledge which often resulted in distortion of the information was more apparent when the text was not available for reference.

IV. CONCLUSION

The present study explored the effect of text availability on performance in EFL reading comprehension tests, and its interaction with the language in which items are constructed and answers should be selected or produced, and with the type of test format used to measure reading comprehension. Based on the results and the discussion presented already, the following conclusions were drawn.

Text availability during test-taking significantly affects reading comprehension test scores. Students obtain much higher scores when texts are available for reference than not. However, the effect of text availability is different depending

on the task type, especially for students of lower proficiency, or depending on the language in which questions are constructed and answers should be selected or produced, i.e., whether it is their native language or the target language. In other words, when the task is very easy or very difficult, students' test scores are not significantly higher when they have test passages at hand than when not. In addition, when students are tested in their native language, it is not a critical factor in test scores whether or not they have test passages for reference.

When the text is not available, students use different reading strategies in answering the questions to compensate for the difficulty caused by the text unavailable condition, but it may not greatly help them get higher scores.

The findings in the present study imply that taking tests without text for reference could be too big a burden for students of low proficiency, especially for the tasks that require production, and if they have to take tests in the target language. It may even discourage them in some cases. For the purpose of reading comprehension research, however, comparing readers' performance when the text is available with that when it is unavailable, may provide valuable information about the processes taking place in the readers' mind when they read and answer the questions.

In addition, the findings in the present study seem to affirm the view that the greater a reader's grammatical ability in the language, the higher the overall comprehension of the texts read (cf. Bernhardt, 1983), again supporting the generally accepted view that syntax can increase text comprehension. Consequently, poor readers are encouraged to expand their vocabulary and to gain greater control over complex syntactic structures in the target language in order to improve reading comprehension in that language. In addition, self-reports by the interviewees of lower proficiency level seem to imply that the lower the proficiency level of readers is, the more interesting the topics of texts might have to be, even in a testing context, so that they could involve themselves in the task with more motivation.

In conclusion, the findings of the present study show that intermediate EFL readers, even those who can be classified as intermediate high, are still affected by the condition of whether or not the text is available for reference during test-taking, and its effect may differ as a function of the other factor(s), as can be shown in its significant interaction with other variables considered in this

study. Therefore, it is important to understand the complexities involved in performance on the (EFL) reading comprehension tests that may result from the influence of many variables, and to consider them in choosing and constructing tests, and in interpreting findings from related research.

REFERENCES

- Bernhardt, E. B. (1983). Three approaches to reading comprehension in intermediate German. *Modern Language Journal*, 67, 111-115.
- Johnston, P. H. (1983). *Reading assessment: A cognitive basis*. Newark, DE: International Reading Association.
- Johnston, P. H. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, 19, 219-239
- Kendall, J. R., Mason, J. M., & Hunter, W. (1980). Which comprehension? Artifacts in the measurement of reading comprehension. *Journal of Educational Research*, 73, 233-236.
- Lee, J. F. (1986). On the use of the recall task to measure second language reading comprehension. *Studies in Second Language Acquisition*, 8, 201-211.
- McLaughlin, G. H. (1969). SMOG Grading-a new readability formula. *Journal of Reading*, 12, 639-636.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 147-179.
- van Dijk, T. A., & Kintsch, W. (1978). Cognitive psychology and discourse: Recalling and summarizing stories. In W. V. Dressler (Ed.), *Current trends in text linguistics* (pp.61-78). New York: De Gruyter.
- Wolf, D. F. (1993). A comparison of assessment tasks used to measure foreign language reading comprehension. *Modern Language Journal*, 77, 473-489.