

English Teaching, Vol. 54, No. 3, Fall 1999

An Analysis of Performance in Two Different Measures of Reading Comprehension : A Preliminary Study

Jeongsoon Joh

(Hoseo University)

Joh, Jeongsoon. (1999) An analysis of performance in two different measures of reading comprehension : A preliminary study. *English Teaching*, 54(3), 239-258.

The purpose of this study was to examine learners' performance in two different measures of reading comprehension: The multiple-choice items task and the summary task. Subjects were 72 Korean high school students. The rate of correct response in the multiple-choice items was compared with that of correct retrieval of corresponding idea units in the summary protocols. Results showed that the students' comprehension of the main idea was not influenced by the type of task while other information in the text was. Probably benefitting from the extra information provided via stems and choices, students generally performed better in the multiple-choice items. Students with lower target language proficiency were more sensitive to the type of task than those with higher proficiency, though the latter also showed some degree of variation depending on the task type. The findings from this study have several implications for English teachers. First, teachers should be careful in constructing multiple-choice items so that students cannot reach an answer simply by using test-taking strategies. Second, the summary task might be considered as a potential measure of reading comprehension as well as a tool for teaching reading skills, with its affluent diagnostic information of students' comprehension of a given text, along with its possible washback effect on the reading behavior of learners.

I. INTRODUCTION

Bransford (1979), in his ever-influential book *Human Cognition*, properly illustrated four basic factors that must be considered in learning, and emphasized the 'interrelatedness' of those factors in explaining and understanding why an individual learner may or may not do well in a particular type of educational environment or experimental task. The four factors he outlined are as follows: the nature of the materials to be learned; the characteristics of the learner; the learning activities; and the criterial tasks, i.e., the kind of test tasks used to evaluate the degree of learning. As he strongly suggests, we should take into account the washback effect of a test as well as the major requirements of a test, such as validity, reliability, and practicality, when deciding the type of task to use as a measure of learning for either educational or research purposes. In other words, we should keep in mind that the impact of an assessment instrument on learning process is of paramount importance.

Recently, one of the major issues in elementary and secondary English education in Korea has been the attempt to develop and administer the "performance assessment" as an alternative to the traditional paper-and-pencil tests, which have the potential danger of measuring pieces of knowledge rather than the actual ability to do something with language. Despite the general agreement on the need for such performance-oriented evaluation, selecting and constructing the type of instruments for the performance assessment is no simple matter. In addition, although several studies have been conducted on this issue for elementary and college level English education (Jeon, 1998; H. Kim, 1998), little research has been conducted with secondary school students. What is more important is the necessity of empirical studies with learners at various ages and English proficiency levels before teachers decide to use a measure for performance assessment. For example, Joh (1997) found that Korean high school students, regardless of their relative target language proficiency, thought that the summary task was the best measure of reading comprehension, without being contaminated by extra reasoning skills such as test-taking strategies or blind guessing. They preferred the summary task as a test format even though they reported the difficulty they experienced while performing the task.

In order to obtain some empirical data concerning this matter, the present study attempted to examine the performance of high school EFL students in two

different measures of reading comprehension, i.e., the multiple-choice items task and the summary task. The former has been most widely used as a measure of reading comprehension, mainly for the relative ease of scoring and achieving high reliability. The latter has rarely been used for the same purpose in Korean secondary schools, but many researchers have acknowledged the value of the summary task not only as a measure of comprehension but also as a means to improve comprehension, when modified and implemented as appropriate for the purpose of the specific learning situation (Brown & Day, 1983; Carrell, 1990; Hidi & Anderson, 1986; Hoyer, 1988; Johns, 1986; S. Kim, 1998; Swaffar et al., 1991; Taylor, 1984).

It is hoped that, by comparing learners' performance in two different measures of quite a different nature, valuable insight would be obtained about the mental processes involved in encoding what readers have comprehended. Moreover, some practical information may be gained about how the difference in performance may relate to the validity of a specific measure. In that sense, the findings from this study are also expected to provide language teachers with insights into assessing reading comprehension, as well as for teaching reading, in their classrooms.

II. METHOD

In this study, the following research questions were sought to answer: 1) Do students perform differently in the multiple-choice items task from when they perform in the summary task?: 2) If they perform differently, what is the nature of the difference?: 3) Is there any difference between higher and lower proficiency students in terms of their performance on different types of task?

1. Subjects

Subjects were 72 high school second graders. They were from the same school but from two different classes. The experiment was conducted just prior to the end of the fall semester. Originally the number of students were 86. However, in the course of classifying students into higher and lower English proficiency groups (in order to compare performance by the two groups), 14 border-liners were excluded from the final analysis. Students' English scores on a mock college scholastic

apitude test administered about a month before this experiment, were used as the criterion for classifying students into two different proficiency groups. Some of the descriptive statistics for the subjects are presented in Table 1. It was assumed that the two different task groups were not different in their proficiency as measured by the pretest and in general characteristics which may affect the result of the experiment. This assumption was based on the fact that subjects in this study were from the same school, so the nature of instruction they received was the same, and they were examined under the same circumstances. The result of t-test also revealed that there was no significant difference in their pretest means (at .05 α level) between the two different task groups ($p = .344$), between the lower proficiency students in the multiple-choice group and those in the summary group ($p = .643$), and between the higher proficiency students in the multiple-choice group and those in the summary group ($p = .654$).

TABLE 1
Mean and Range on the Pretest

	Mean	Range
Total Sample (N=72)	70.2	20.6-100.0
Low (n=33)	50.5	20.6- 68.8
High (n=39)	88.8	78.8-100.0
Multiple-choice (n=38)	68.9	20.6-100.0
Low (n=19)	49.4	20.6- 68.8
High (n=19)	88.3	78.8-100.0
Summary (n=34)	71.7	21.9- 96.3
Low (n=14)	51.9	21.9- 68.1
High (n=20)	89.2	78.8- 96.3

2. Materials

An expository passage from a reading textbook was used as the test text. According to the authors, both native speakers of English, it was intended for ESL/EFL students at the intermediate and high intermediate level in high school, pre-college, or college programs, as well as for native speaker developmental education classes. The passage consists of 429 words. Four high school teachers confirmed that the vocabulary and grammatical structures in this passage did not exceed those covered in the subjects' curriculum. The degree of difficulty of this passage calculated using the readability formula developed by McLaughlin (1969)

was 9, which means that it is appropriate for 9th graders in American school system. As for the rhetorical organization, the passage consisted of a list of causes of a single effect (i.e., longer life expectancy of women) and explanation of the causes.

A model summary of the passage was produced in the following manner. First, I asked three native speakers of English majoring in TESL to summarize the passage, then I compared the three different summaries, abstracted those idea units that were included in all three summaries, and created a new summary. This new summary was examined by another native speaker of English majoring in TESL to ensure its natural quality. This model summary was used as the criterion for scoring the summary protocols produced by the subjects, as well as the basis for constructing multiple-choice items.

Five multiple-choice items were constructed¹⁾ so that those five questions would cover the major propositions of the model summary. In constructing items, special caution was exercised so that the subjects would not get extra information from a stem and/or choices which may have given apparent clues to answer a different question.

3. Procedure

Subjects were randomly assigned either to the multiple-choice items task or to the summary task. They took these tasks during their regular class hour. Subjects were given a short instruction about how to produce a summary not a verbatim recall.²⁾ They were not allowed to refer back to the passage while working on the

1) In this study, subjects were instructed to summarize in Korean, their native language. This was based on the findings from previous research that summarizing in the target language was too difficult a task for average Korean high school students (cf. Joh 1997). Accordingly, multiple-choice items were also constructed in Korean so that performance of subjects in two different types of task could be compared.

2) The instruction was as follows: Summarize what you read by using 120-150 words. Be sure to synthesize well what you read: abstract the main idea and important supporting ideas, 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 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 express them (in your own language if possible).

task, and were informed of this condition before they began. This condition was implemented based on the arguments and practices in previous studies (Johnston, 1984; Kendall et al., 1980; Wolf, 1993): in order to prevent the reader from utilizing test-taking strategies, such as matching elements in test items with those in the text. It is generally believed that test-taking strategies may weaken the validity of a test by, for example, measuring extra reasoning skills which is not necessarily the object of assessment. Subjects were given 25 minutes at maximum, based on the time estimation obtained from a pilot study.

4. Analysis of data

Data were analyzed in the following manner.

- 1) The summary protocols were examined in terms of five core semantic units, which were mainly in the form of proposition.³⁾ These five semantic units were selected so that they should correspond to the five items in the multiple-choice format.
- 2) Students' answers in the multiple-choice items were analyzed to see if there was any relationship between what they chose as answers and what they wrote in the summary protocols.
- 3) For the purpose of exact comparison with the performance in the multiple-choice items, the summary protocols were scored either correct or incorrect, depending on whether each statement in the summary protocol fully covered the proposition (or phrase) which was the correct response for each question in the multiple-choice items. The rate of correct response for each question in the multiple-choice task was compared with the rate of the correct proposition in the summary protocol corresponding to the correct response for the multiple-choice item.

Examples of scoring guidelines for the summary task are presented in Table 2.⁴⁾

3) In the case that the correct response to a question in the multiple-choice items was a phrase, not a clause, the response in the summary task corresponding to that specific multiple-choice item was examined with the same criterion, for the sake of comparability.

4) In the example given in Table 2, other answers were also counted correct if they mentioned the word 'reason' even though it was not included in the same sentence

TABLE 2
Examples of Scoring Guidelines for Summary Task

Multiple-choice	Summary
Question 1. What is the best topic of the given passage?	Corresponding propositions considered correct:
Correct choice: c. The reason that women generally live longer than men.	"Women (generally) live longer than men for (several/two/biological and cultural) reasons" "This passage is about the reasons that women live longer"

III. RESULTS AND DISCUSSION

Table 3 shows the result of comparison between performance in each of the multiple-choice items and performance in each of the propositions in the summary protocols corresponding to each of the multiple-choice items.⁵⁾

TABLE 3
Comparison of Performance (Rate of Correct Response) in Two Measures

	Multiple-Choice Group	Summary Group
Part I	Low : 86%	Low : 86%
: Main Idea	High : 100%	High : 100%
	Total : 92%	Total : 84%
Part II		
: Two factors in women's longer life	Low : 74%	Low : 43%
	High : 84%	High : 65%
	Total : 79%	Total : 56%
Part III		
: Function of a biological factor	Low : 79%	Low : 29%
	High : 95%	High : 55%
	Total : 87%	Total : 44%

as the one containing the main idea. For example, some students wrote, "Women (generally) live longer than men. The reason for this is...", "Women live longer than men. There are several [two] reasons for this," or "Women live longer than men. This is because..." These were evaluated as correct.

- 5) In Part II of the summary task, only those responses that explicitly mentioned both of the two factors ('biological and cultural factors') were counted as correct, while those that mentioned only examples of either biological or cultural factors (such as female hormones, female genes, drinking, smoking, stress) were counted as incorrect.

Part IV		
: Function of a cultural factor	Low : 11%	Low : 7%
	High : 53%	High : 55%
	Total : 32%	Total : 35%
Part V		
: Prognosis of future and its basis	Low : 53%	Low : 21%
	High : 74%	High : 75%
	Total : 63%	Total : 53%

Note 1. 'Low' and 'High' represent the low and high proficiency group students participating in this study.

2. Numbers were rounded.

In order to get some information, albeit indirectly,⁶⁾ about readers' mental processes while they were selecting a specific choice as their answer to each question in the multiple-choice items, subjects' responses were analyzed in terms of the frequency with which each choice was selected. The result of this analysis is presented in Table 4.

As shown in Table 4, in the rate of correct response to the main idea of the given text, there was practically no difference either between different tasks or between different proficiency groups within the same task group. One thing to note is that while all the higher proficiency group students were correct in choosing or expressing the main idea, some of lower proficiency group students did not show their full understanding of the main idea. A closer look at their responses revealed that in the case of the multiple-choice items, students selected the choice which was inappropriate as a correct response because it expressed only part of the necessary semantic elements. In the case of the summary task, all the responses that were not given credit for the main idea, lacked the crucial element of 'cause' or 'reason' for women's longevity: they simply wrote down, "Women live longer than men." or did not mention anything about it in an appropriate position in their summaries.

This result shows that although most students were aware of, and able to express, the main idea, some students of limited proficiency had difficulty in abstracting the gist of a given text, and therefore could be distracted by partial

6) In this study no introspective or retrospective self-report was collected. Therefore any statement made about readers' mental processes in selecting a specific choice would be indirect reasoning based on the results of response analysis presented in Table 4.

expression of the main idea, or by substitution of subordinate term for superordinate term (e.g., '[the influence of] *hormones* and *stress*' for *biological* and *cultural* factor'). This finding also implies the need for the teacher to train lower proficiency students in distinguishing parts from the whole, and key points from supporting ideas or examples.

With regard to the two factors affecting women's longevity, there was considerable difference in the rate of correct response between the multiple-choice group and the summary group. The differences between the two proficiency groups were larger within the summary group than within the multiple-choice group. In the multiple-choice group, although the majority of the responses were correct, there were still some students confused by distracters such as 'physical and mental' or 'psychological and biological', where the correct response was 'biological and cultural' factor. These students seemed to resort to their general knowledge about the world when they could not remember the exact terms as they appeared in the text.

TABLE 4
Response Analysis of Multiple-Choice Items

Stem and choices for each item	Frequency of each choice selected	
	High	Low
Q1. What is the main topic of the given text?		
a. Cultural factors that help women live long	0	0
b. Biological reasons that women get sick less frequently	0	1
*c. The reason that women generally live longer than men	19	16
d. How hormones and stress influence life expectancy	0	2
Q2. What are the two major factors that influence women's longer lives?		
a. sociological and cultural factors	0	0
*b. cultural and biological factors	16	14
c. physical and mental factors	2	3
d. psychological and biological factors	1	2
Q3. What is it that helps to regulate blood pressure and resist infections?		
*a. female hormones	18**	15
b. female genes	0	1
c. women's body cells	0	1
d. women's blood flow	0	2

Q4. According to the author, which is an example of the factors that can be related to life expectancy?		
a. the number of hours you work a day	8	
b. the kind of cigarettes that you smoke	0	
*c. the position that you hold in your office	10	
d. physical condition of your home	1	
Q5. Does the author think that the situation is likely to change?		
a. Yes. Because the cultural reasons for women's longer life are changing.	4	5
*b. No. Because the biological reasons for women's longer life will not change.	14	10
c. Yes. Because the reasons for women's longer life are not not true any more.	1	1
d. No. Because cultural and biological reasons will keep women living longer.	0	3

Note 1. The asterisk represents the correct answer for each question.

2.**One student in the High Group did not response to this question.

Examination of students' responses in the summary protocols clearly showed how the extra information provided by the choices in the multiple-choice format could give the reader retrieval cues needed to answer a given question. As shown in Table 3, about 80% of students were correct in the multiple-choice format, whereas only about 50% were correct in the summary format which did not provide any extra information for the reader. Lower proficiency students benefitted much more by this than their higher proficiency counterparts.

Lower proficiency students had a general tendency to list examples of each factor without mentioning the superordinate term that can cover those individual examples. On the other hand, those higher proficiency students who did not get full credit for their response, generally mentioned one of the two factors along with the examples of the other factor, e.g., 'hormones and cultural factors (such as drinking and smoking...)', or 'biological factors and stress, drinking, smoking.....' This implies that lower proficiency students generally have more difficulty in categorizing elements into a superordinate concept. And this seems to be another area that the teacher should take into account in teaching reading.

With regard to the function of a biological factor (Part III), the difference in the rate of correct response between different task groups was the largest: 87%

for the multiple-choice group, and 44% for the summary group. The difference between the two proficiency groups was much smaller in the multiple-choice task than in the summary task, which did not provide any retrieval cues and therefore had a more detrimental effect on the lower proficiency students. In the multiple-choice group, all higher proficiency students selected the correct answer, while a few lower proficiency students were distracted by the other biological factors mentioned somewhere in the text. Two of them selected 'blood flow' probably because it resembled one of the words in the stem - 'blood pressure'. This exemplifies how some students, especially lower proficiency ones, utilize what is called matching strategy while they work on multiple-choice items. Such a strategy sometimes works, but not always as this case shows. What the teacher should keep in mind when constructing multiple-choice items, is to make sure that this type of test-taking strategy will not work very well.

In the summary task, about 30% of lower proficiency students and about 55% of higher proficiency students correctly mentioned both 'female hormones' and their function in relation to women's longevity. This rate is very low compared with that obtained in the multiple-choice task, which was 79% and 95%, respectively. Although most of the students understood, and succeeded in selecting, the function of female hormones in the multiple-choice task, they were not successful in explicitly mentioning it in their summary. Among those responses that did not get full credit for this part, some mentioned 'female hormones' but not their functions, and others were inadequate or vague about their functions.

Generally, lower proficiency students either did not ever mention 'female hormones', or did not explain at all how they help women's longevity even when they mentioned it. On the other hand, those responses by higher proficiency students that did not get full credit were vague about 'functions' though all of them mentioned 'hormones' and most of them even tried to explain how hormones related to women's longevity. Students may have not felt the need to describe the 'functions' of female hormones once they mentioned hormones as an example of biological factors. This reasoning could be applied to higher proficiency students. On the other hand, it is possible that they, especially lower proficiency students, either did not fully comprehend the specific part of the text due to limited knowledge of the target language, or did not have enough time (or attention capacity) to store the information into their long-term memory for retrieval, even though they understood while they were reading. This is what usually happens

with lower proficiency students. Therefore, as many researchers have suggested, the teacher needs to encourage poor readers 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 (Bernhardt, 1983; Cooper, 1984).

As shown in Table 3, the rate of correct response was lowest in the fourth part both in the multiple-choice items and summary task, in which students were required to select, or describe, an example of cultural factors that could influence human longevity. The given text described three major cultural factors - drinking, smoking, and stress, and stress-related differences between men and women were explained more in detail than drinking or smoking, probably because it is so well known that men generally drink and smoke more than women. And the fourth question in the multiple-choice format required exact comprehension of how a certain (cultural) condition might affect life expectancy. As shown in Table 4, only 2 out of 19 in the lower proficiency group selected the correct choice in the multiple-choice items task. The majority of students selected 'the number of hours you work a day' or 'the physical condition of home'. Two of them even chose 'the kind of cigarette you smoke'. In the case of higher proficiency students, 10 out of 19 recognized the correct response: 8 chose 'the number of hours you work a day', and one selected 'the physical condition of home', but no one in this group was distracted by 'the kind of cigarette you smoke'. The number of hours of daily work was not mentioned in the text, but students seemed to depend on their general knowledge again. They may not have retrieved the information required to answer this question, either because of their inadequate comprehension of that specific part of the text or because of constraints on memory. The extra information available in the stem and choices was not very helpful in this case, because the choices presented were not verbatim repetition of phrases in the text. Instead, students had to logically restructure the information given in the text in order to recognize the correct answer among distracters. In other words, this question required some inferring skill.

This seemed to impose much more difficulty on the lower proficiency students, for they showed almost the 'floor effect' on this question. Nearly half of low proficiency students were distracted by the incorrect choices that contained such words as 'smoke' and 'physical condition', which appeared in the text. In this case their matching strategy failed once again. This finding implies that an item that

requires inferring skill could impose a great burden on the readers, especially those with limited target language proficiency. When the reader meets phrases that are not exactly the same as those in the text, s/he may get easily confused and this requires more intense analysis to ascertain whether they are semantically equivalent or not. This phenomenon may be more prevalent among learners of lower target language proficiency. Therefore, language teachers need to help their students improve the skill of inferring information that is not explicitly stated in the text.

On the other hand, in the summary task, most students mentioned smoking or drinking as a factor which can shorten life expectancy. However, only half of the lower proficiency students mentioned stress, and only one of them correctly mentioned in what manner the stress factor influences women's longer lives. As for the higher proficiency group, all mentioned stress as a factor in life expectancy, and 55% of them correctly described how it relates to women's relatively longer lives than men. The other 45% were vague, or too general, in describing this point. This finding implies that the lower proficiency students tend to remember the key words such as hormones, genes, drinking, smoking, stress, but have difficulty in subsuming these elements into a superordinate category, as well as in associating a key word with the information related to that key word. Once again, this could be ascribed to their limited target language proficiency, when we see that the higher proficiency students in this study were much more successful in these skills.

The last part to be examined concerned the possibility that women's longevity will continue. Both the prognosis and its basis were required in the response for it to be considered correct. While higher proficiency students were the same in the multiple-choice and summary tasks (75% of correct response in both tasks), lower proficiency students did show much difference depending on the type of task (53% of correctness in the multiple-choice, but 21% in the summary task), implying that they benefitted again from the retrieval cues provided by the information in the choices. In the multiple-choice task, those who selected incorrect choices were mostly distracted by the choice, "Situations will change (i.e., Women will no longer live longer than men) because the cultural reasons for women's longer life are changing." Even several higher proficiency students selected this as their response. This statement is partly true, but not false on the whole, because the text explicitly stated that women will keep living longer than men due to the *unchanging biological factors* even though the influence of cultural factors is

becoming less powerful in the course of social changes. This finding implies once again that teachers need to train students to distinguish the part from the whole, an example from the higher category that can subsume the example.

In the summary task, half of the lower proficiency students stated that women would keep living longer than men, but only 21% of them mentioned the basis for such prognosis. On the other hand, 75% of the higher proficiency students correctly mentioned both of the key points, whereas the rest either did not mention this last major part of information or did not explicitly mention the basis for their correct prognosis. The lower proficiency students in this study may not have recognized the importance of the specific proposition in constructing a summary, or not have comprehended enough to retrieve the information at the time of summarizing. The general benefit called 'recency effect'-that the pieces of information presented most recently are more likely to be learned than those in the middle-did not seem to help greatly the lower proficiency students in this case. Probably their limited command of the target language and capacity constraints might have superseded the potential benefit of the recency effect.⁷⁾

IV. CONCLUSION

In this study, learners' performance in two different measures of reading comprehension was analyzed and compared. The results showed that students perform differently depending on the type of task used to measure their reading comprehension, and the difference in performance varied depending on the kind of information given in the text. For example, students adequately comprehended and remembered the main idea in both tasks, while they revealed moderate to dramatic difference in understanding and retrieving some crucial pieces of information in the text, depending on the type of task. In most cases, the information provided via stems and choices in the multiple-choice items seemed to contribute to better performance in the multiple-choice task. Simple guessing or matching strategy may have played a role, too. Students with lower target language proficiency were

7) The proposition in question was located near the end of the text, and naturally would be located at the end of the summary protocol. Therefore, as they approached the latter parts of the text while reading, and as they got involved longer in a relatively difficult task, their concentration/attention capacity may have decreased.

generally more affected by the type of task, though higher proficiency students also showed some degree of fluctuation in the two different measures. That is, lower proficiency students performed much better in the multiple-choice items (almost twice the rate of correctness), except when they showed a ceiling or floor effect. They depended more on general knowledge or a simple matching strategy when they were not able to retrieve what they read. This sometimes resulted in distortion of the original information in the text (in the case of the summary task), or led them to be distracted by partly true statements (in the case of the multiple-choice task).

Based on the results and discussion presented above, the following conclusions and implications could be drawn. The summary task can be used as a potential measure of reading comprehension in that it gives more clear information about the reader's comprehension of the text than the multiple-choice task which cannot tell, without self-report of the reader, how the reader has reached the specific answer s/he eventually chose. As the result of this study shows, students often benefitted from the extra information and the consequent retrieval cues given via stems and/or choices, while in the summary task they could not retrieve that information. Some people have questioned the validity of the summary task as a measure of reading comprehension in that it requires production which could be of quite different nature from comprehension. This cannot be denied. At the same time, however, achieving the construct validity in a strict sense could not, and probably should not, counteract all the other educational considerations, such as the possible influence a test may have on the following learning processes and importance of providing learners with opportunities to practice useful language skills. Actually, some reading researchers included, as an element in defining "reading," the ability to transfer what was understood from written materials into other modalities of language, e.g., speech or writing.⁸⁾ In addition, the potential threat to validity that comes from production skills intervening in comprehension tasks could be mediated, at least to an extent, by using the native language of the learners as the language of production: that is, by not forcing production in the target language. This view is partly based on the assumption that average learners reaching the age of 16 or

8) This was a view presented in a graduate course in TEFL methodology by Professor Bordie at the University of Texas at Austin. Unfortunately, it is not quite certain if it was his personal view or if there had been several scholars who shared this view. However, I personally agree to this view, especially when it comes to the "reading" within educational contexts.

17, like those in this study, are generally considered to have almost perfect command of their native language and some cognitive, metalinguistic skills required to express their ideas in written language.

The summary task is also useful as a provider of diagnostic information. It clearly shows what the learner comprehended as well as what s/he did not comprehend about the text. Thus the teacher can find the weak points of the individual learner, and help the learner improve specific reading skills s/he lacks. Summary protocols also reveal what was going on in readers' minds while reading, by way of addition, omission, or distortion of the content of the original text. It may come from various sources: the reader's use of prior knowledge, or his/her ignorance of specific vocabulary required to express understanding of the text, among others. This kind of information may provide valuable information to researchers of second language reading, as well as to language teachers. Therefore, the summary task without text for reference while summarizing, can be a useful tool for investigating reading in a second language, as well as for educational purposes.

The findings from this study seem to suggest several implications for teachers and/or test constructors. That is, if the multiple-choice format is to be used mainly for economy's sake, special caution should be exercised so that the items themselves will not give clues to the answers; otherwise, students may arrive at the correct answer by utilizing a simple matching strategy or by depending on general knowledge. This may weaken the validity of the items.

Where a close investigation of a reader's comprehension of a given text is of primary concern, the summary task can give a full picture of his/her understanding of a main idea, ability to distinguish key concepts from supporting ideas, and superordinate categories from subordinate examples of those higher order terms. It can also show if the reader has the ability to organize the information s/he has obtained from the given text into coherent discourse, which is an essential skill in an academic environment. Analysis of the performance in the summary task could provide teachers with valuable diagnostic information of a learner's reading comprehension, so that they could make specific plans for teaching the reader those skills that s/he needs most.

The summary task is also to be implemented as an instrument for "performance assessment." The ability to properly summarize a text is surely a measure of performance (or performing ability), not of discrete pieces of knowledge. Moreover,

in terms of integration of skills, using the summary task as a post-reading activity could be a useful teaching strategy as well as helpful learning strategy. And this will hopefully encourage meaningful, holistic reading on the part of the learner and will provide opportunities to practice writing skills at the same time. Swaffar et al. (1991) properly grasped this point by suggesting that reading comprehension tests should connect a student's comprehension of the text to the *expression* of the ideas that the reader has about what has been comprehended. Consequently, they argued, tests of reading must reveal that students are able to comprehend *and* utilize or express views about that meaning. According to Swaffar et al., one of the test's goal should be to elicit either L1 or L2 discourse that expresses a reader's views of textual implications. The summary task seems to be an appropriate means to serve that goal. For instance, Hoyer (1988), who examined the effectiveness of the summary task as a reading comprehension strategy, found a positive correlation between summary writing ability and reading comprehension ability.

In the present study the summary task was performed in the students' native language considering their target language proficiency. However, with learners of higher proficiency, summarizing in the target language could be used as a post-reading activity and/or as a measure of reading comprehension.

REFERENCES

- Bernhardt, E. B. (1983). Three approaches to reading comprehension in intermediate German. *Modern Language Journal*, 67, 111-115.
- Bransford, J. D. (1979). *Human Cognition*. Belmont, CA: Wadsworth, 3-11.
- Brown, A., & Day, J. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Memory and Language*, 22, 1-14.
- Carrell, P. L. (1990). Training formal schemata—replication results. In L. A. Arena (Ed.), *Language proficiency* (pp. 85-92). New York: Plenum.
- Cooper, M. (1984). Linguistic competence of practised and unpractised non-mature readers of English. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 122-135). New York: Longman.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications of instruction. *Review of Educational Research*, 56, 473-493.

- Hoye, M. (1988). *The effects of summary writing on the reading comprehension of American and ESL university freshmen*. Unpublished doctoral dissertation. University of Texas at Austin.
- Jeon, J. (1998). Performance assessment: Use of self-assessment (SA) for the communicative college English. *English Teaching*, 53(3), 179-198.
- Joh, J. (1997). *The Effects of assessment features on performance in EFL reading comprehension tests*. Seoul: Hanshin.
- Johns, A. M. (1986). Coherence and academic writing: Some definitions and suggestions for teaching. *TESOL Quarterly*, 20, 247-265.
- 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.
- Kim, H. (1998). Chotung yenge kyoywuk-ui swuhayng p'yenkga panghyang tamsayk [Performance assessment in elementary school English education: Approaches and techniques]. *English Teaching*, 53(4), 223-245.
- Kim, S. (1998). An analysis of Korean EFL readers' summarization of L2 texts. *English Teaching*, 53(4), 63-91.
- McLaughlin, G. H. (1969). SMOG grading - a new readability formula. *Journal of Reading*, 12, 639-646.
- Swaffar, J. K., Arens, K. M., & Byrnes, H. (1991). *Reading for meaning: An integrated approach to language learning*. Englewood Cliffs, NJ: Prentice Hall.
- Taylor, K. K. (1984). The different summary skills of inexperienced and professional writers. *Journal of Reading*, 27, 691-699.
- Wolf, D. F. (1993). A comparison of assessment tasks used to measure foreign language reading comprehension. *Modern Language Journal*, 77, 473-489.

APPENDIX A

The Text Used

Both men and women are living longer these days in the industrialized countries. However, women, on the average, live longer. In general, they can expect to live six or seven years more than men. The reasons for this are both biological and cultural.

One important biological factor that helps women live longer is the difference in hormones between men and women. Hormones are chemicals which are produced by the body to control various body functions. Between the ages of 12 and 50, women produce hormones that are involved in fertility. These hormones also have a positive effect on the heart and the blood flow. In fact, women are less likely to have high blood pressure or to die from heart attacks. The female hormones also protect the body in another way. They help the body to defend itself against some kinds of infections. This means that women generally get sick less often and less seriously than men. The common cold is a good example: women, on average, get fewer colds than men. Women are also helped by their female genes. Scientists are still not exactly sure how genes influence aging, but they believe that they do. Some think that a woman's body cells have a tendency to age more slowly than a man's. Others think that a man's body cells have a tendency to age more quickly. Recent research seems to support both of these possibilities.

The cultural context can also influence expectancy for men and women. (Life expectancy is the expected length of a person's life.) For example, women generally smoke cigarettes less than men. They also drink less alcohol on average. Both cigarettes and alcohol have been proven to cause many health problems and to shorten lives. Another factor that has influenced the lives of women is the lack of stress. Stress is well known to shorten lives. Until recently, women who worked were usually in less responsible, less stressful positions. At home, housework tends to keep women in better physical condition than men. This generally better physical condition is yet another factor in women's longer lives.

These cultural factors have played an important part for women who are now getting old. But the social habits of women are changing. Young women are smoking and drinking more than women used to. More women are working now and holding more responsible positions. These changes may mean that the cultural context will no longer help women live healthier lives. However, the other, biological factor in life expectancy remain unchanged. Women probably will therefore continue to live longer than men.

APPENDIX B

A Model Summary of the Text

Although both men and women are living longer these days, women, on the average, live longer than men for both biological and cultural reasons. Biologically, female hormones not only have a positive effect on the heart and blood flow but also protect the body against some sorts of infections. In addition, according to recent research, female body cells tend to age more slowly than those of men. Another factor that affects life expectancy is the cultural context, such as smoking, drinking, and amount of stress. Women have been in generally better physical condition than men and so live longer because they tend to smoke less, drink less, and work usually in less stressful positions. However, due to the changes in social habits,

these cultural factors may no longer help today's women lead healthier lives. Nevertheless, due to the unchanging biological factor, women will still enjoy a longer life expectancy than men.

APPENDIX C

Multiple-Choice Items

1. 이 글은 주로 무엇에 관한 것인가?
 - a) 여성이 오래 살도록 하는 문화적 요인
 - b) 여성이 병에 덜 걸리는 생물학적 이유
 - c) 여성이 일반적으로 남성보다 오래 사는 이유
 - d) 호르몬과 스트레스가 수명에 미치는 영향

2. 여성의 장수에 영향을 미치는 두 가지 중요 인자는 무엇인가?
 - a) 사회적, 문화적 요인
 - b) 문화적, 생물학적인 요인
 - c) 신체적, 정신적 요인
 - d) 심리적, 생물학적 요인

3. 혈압을 조절하고 감염에 대한 저항력을 가지도록 도와주는 것은 무엇인가?
 - a) 여성 호르몬
 - b) 여성의 유전자
 - c) 여성의 신체세포
 - d) 여성의 혈류

4. 윗 글에 따르면, 수명과 관련있는 요인의 예가 될 수 있는 것은 무엇인가?
 - a) 하루에 일하는 시간의 수
 - b) 피우는 담배의 종류
 - c) 회사에서의 직책
 - d) 집안의 물리적 조건

5. 윗 글의 저자는 상황이 변화하리라고 생각하는가? 그리고 그 이유는 무엇인가?
 - a) 그렇다. 왜냐하면 여성이 더 오래 살게 하는 문화적 요인들이 변화하고 있기 때문이다.
 - b) 아니다. 왜냐하면 여성이 더 오래 살게 하는 생물학적 요인들이 변하지 않을 것이기 때문이다.
 - c) 그렇다. 왜냐하면 여성이 더 오래 살게 하는 요인들이 더 이상 적용되지 않기 때문이다.
 - d) 아니다. 왜냐하면 사회적 요인과 생물학적 요인이 계속 여성의 장수에 영향을 줄 것이기 때문이다.