The Effects of Differing Language Test Presentation Methods on EFL College Students' Listening Test Performance

Jinseon Koo and Eun-Joo Lee[*]

This study investigates the effects that three different test presentation formats, i.e., written Korean, written English and aural English, have on EFL learners' listening comprehension. A total of 169 Korean college students participated in this study. The participants were divided into six groups, by three presentation format types and two proficiency levels. They were presented with multiple-choice English listening tests. The results of the study showed that listening questions using the written Korean presentation mode were the easiest for them, while the aural English options were the most challenging. The analysis showed that high and low proficiency level students performed differently on the listening tests. The high level students performed best in written Korean and written English test formats, and had the most trouble with aural English tests. Students of both proficiency levels performed measurably better with test options written in Korean as compared to the other two presentation modes. Results of the post-test survey showed that test options written in participants' L1 were most preferred by the students, while the aural English option was the least favorable test method. However, the participants considered the aural English test mode to be the most fair and appropriate method of assessing test-takers' listening comprehension abilities.

Keywords: English listening test, presentation mode of listening question, aural English options, language of options

1 Introduction

As part of measuring students' overall English abilities, assessing listening skills is becoming a more significant component of language testing. In Korea, the College Scholastic Ability Test (CSAT) and the Test of English

---

[*]First author: **Jinseon Koo**, Ph.D. Candidate, Ewha Womans University
Corresponding author: **Eun-Joo Lee**, Professor, Department of English Education, Ewha Womans University

for International Communication (TOEIC) are widely implemented as means to measure students' listening abilities. The listening comprehension questions in CSAT were adopted in 1994 in Korea. The Korean 2021 CSAT includes 17 listening questions out of 45 items, which represents almost 40 percent of the total questions given, while the remaining test question items are primarily reading comprehension items[1]. A widely-used standardized English test in Korea, TOEIC, is comprised of a listening section and a reading section. The listening section forms half of all the test questions.

Several questions are raised regarding how best to assess test-takers' listening skills (Brunfaut, & Revez, 2015; Hansen & Jensen, 1994; Rost, 2016; Rukthong & Brunfaut, 2020; Taylor & Geranpayeh, 2011). One of these questions is whether a student's ability to understand a test question is affected by the test method used. The effects of the 'presentation of options,' which a teacher faces when deciding how to phrase or present a question to the students may affect both their performance and their comprehension. If so, are the differences found in the varying methods in which a question is presented, a matter of importance in helping to either improve or grade students' listening performance? It seems necessary that test developers and language teachers both notice and give some consideration to these particular factors in listening tests, as they can affect test validity.

Taking these characteristics into account, the present study examines how listening test results vary when test-takers are tested using different types of testing methods. This study also investigates how the testing methods used influence test-takers' performance with regards to students' English proficiency levels. Additionally, how a students' perception differs towards different test formats is also valuable information in evaluating their listening skills more accurately. The results of this study may contribute to the development of listening tests which more accurately measure students' L2 proficiency. The advancement and promotion of the design of effective listening comprehension teaching methods can then be applied in the L2 classroom setting and provide greater insight into English education as a whole.

2 Literature Review

Many different types of test methods or formats have been implemented in language testing, and one of the important issues is the effect that test formats themselves have on test performance (Bachman & Palmer, 1996). A number

---

[1] Korea Institute for Curriculum and Education (2020)

of studies have shown how different test methods and formats influence students' language test performance (Filipi, 2012; Hwang & Lee, 2020; In'nami & Koizumi, 2009; Lee & Winke, 2013; Lin & Chen, 2020; Park & Jung, 2019; Rahimi, 2007; Shohamy, 1984). The language of questions and options also has long been discussed in educational research.

Park and Jang (2019) investigated how the order that information is presented in listening and reading tests can affect test performance. They compared two factors in both listening and reading test situations. First, they studied how the differences in the length of a given test answer option affected test scores. Second, they studied whether test scores fluctuated based on the order that information was presented in the given test questions. They found that when the order of listening passages, rather than reading passages, was given in the same order as the answer options, test results were much better than when the presentation answer order was given in a different order.

Hwang and Lee (2020) looked into text variables which could affect test scores and studied the relationship between text syntactic complexity and test scores. They found that some linguistic devices such as syntactic similarity and frequencies of negative expressions affected test performance. Test questions that reflected the same or similar wording used in the passages given, and that also minimized the spacing between new words, resulted in better test performance. The use of negatives within a test question resulted in lower test performance. These results could prove to be helpful in continuing to create fairer and more accurate reading tests.

The ALC (Assessment of Language Competence) listening test, which is a general language proficiency test in six foreign languages, was utilized by Filipi (2012) in an interesting listening test case study. She attempted to explore differences in performance when test questions were given using a student's target language, as compared to tests given in English. She concluded that the questions in the target language were more challenging and difficult. She also found that students' answers were influenced not only by the language variable, but by any additional general understanding or inferences students were able, or required, to infer from the questions themselves. Furthermore, if the questions were explicitly and directly presented, students were more successful in responding correctly. Results of this study showed that simple and direct questions, where there was little competing information to consider, were answered more successfully.

Yi'an (1998) dealt with the issue of how Chinese EFL students employed linguistic and non-linguistic knowledge in completing a multiple-choice question format exam by adopting an immediate retrospection research methodology. The results of the study identified that the test-takers constantly used their non-linguistic knowledge with the linguistic cues from the options provided. The study showed that multiple-choice test formats permitted much uninformed guessing and it led the listeners to depend on their non-linguistic knowledge. These types of options occasionally allowed the students to choose a correct answer, but not for the appropriate reasons.

Additionally, it was found that previewing multiple answer choices helped facilitate the understanding of more advanced students, whereas it seemed to only add to the difficulty for less proficient students.

An inquiry by Lee and Winke (2013) investigated whether the number of possible answers presented as options affected the difficulty of the question asked. They conducted the CSAT (College Scholastic Ability Test) of English listening in their study, which employs three different question formats: with three, four and five answer options. They found that question formats with three answer options were the easiest and question formats with five options were the most challenging, but the results were not statistically significant. More than half of the test-takers preferred the question format with three options because they found them to be easier. The educational policy-makers and test developers, however, seemed to favor five options for high-stakes tests such as the CSAT. These findings also contradicted the results of other studies that showed that a greater number of options produced higher test reliability.

Shohamy (1984) administered a reading comprehension test with 12th graders in an EFL setting in Israel. She investigated if, and to what extent, different test methods measuring reading comprehension influenced test-takers' performance. Consequently, she found that students' reading comprehension scores were affected by using different testing formats such as multiple-choice and open-ended formats in the target language and the test-takers' L1. This study revealed that open-ended formats were more difficult than multiple-choice formats. She also reported that the English version was always more challenging than the L1 version. Questions written in the students' L1 were found to have a lower difficulty level because of the reduction of foreignness of distractors, and additionally, most likely because of a better understanding of L1 questions. She found that lower-level students were more sensitive, and were influenced to a greater degree, by the test method variables as compared to the higher proficiency learners.

In a study involving Iranian college students in Persia, Rahimi (2007) administered L2 multiple choice reading comprehension tests. He examined how language presentation in test questions influenced reading performance on Persian native language students who majored in English. He showed that the test method, as a whole, made no difference on the performance of the test-takers. He also reported that test-takers L1 comprehension abilities were not relevant to their reading comprehension test scores. However, his results partly reinforced those of Shohamy (1984), which showed that lower proficiency students were more sensitive to the language presentation of test items. Essentially, low-level students achieved significantly better performances when the questions were written in their L1.

A comparison of multiple-choice and open-ended formats in listening and reading test performance by using meta-analysis was completed by In'nami and Koizumi (2009). They adopted the use of meta-analysis to

quantitatively analyze test format effects on test-takers' performances. The results suggested that open-ended formats were more difficult for students as compared to multiple-choice formats. They also compared the relative difficulty of the two formats. Their results indicated that the differences in the degree of test format effect ranged from medium to large in L2 listening.

In summary, the studies on reading and listening test options have shown mixed results. Some studies (Filipi, 2012; Shohamy, 1984) have found that options written in a target language are more difficult than answer options given in the test-takers' first language, while other studies (Rahimi, 2007) showed that there was not a significant difference shown between language presentation options. Additional research is clearly needed, to further explore how different test formats affect students' performance, while simultaneously considering the effect that different proficiency levels have on test performance in a variety of test modes. Exploring what test-takers preferences are regarding the language of question presentation options is also a worthy path of inquiry. At this time it seems that a study using a multiple-choice listening test, with the three different test method variables of written Korean, written English, and aural English, has yet to be conducted. Research studying more fully whether or not a student's performance is influenced by the type of testing options available to teachers may prove to be essential in the future formation of language skills tests.

2.1 Research questions

The purpose of this study is to examine the influence that some different formats of test presentation have on performance when used in an English listening test. This study investigated how the effects of the language used varied according to a student's English level. In order to achieve the aims of the study, answers to the following research questions were sought:

1. Do test-takers perform differently when different test formats are used to measure listening comprehension? If so, in which test formats do they perform better?

2. Do different test presentation formats have a different/similar effect on performance, and is this related to test-takers' language proficiency levels?

3. What are the participants' preferences and attitudes towards the three different types of test options?

3 Method

3.1 Participants

The participants in the study were 169 college students in Seoul and Gyeonggi province, who majored in either early childhood education, food and nutrition or computer science. Their grades varied from freshman to junior and the average student age in this study was 20.13 (Minimum = 17, Maximum= 39). 44 students were male (26%) and 125 students were female (74%). None of the participants have previously lived in a country where English is used as a main language. Participants were low- or mid-level students, TOEIC scores range from 300-600. Generally, these students have been learning English for about 10 years at public schools, as English is a mandatory subject from the third grade in elementary school in Korea. These particular students, however, have had few opportunities to speak English outside the classroom or in other EFL learning environments. The research was conducted during their class time.

3.2 Materials

In order to investigate how different test formats might influence students' performance, 30 questions were chosen from the TOEIC Actual Test textbook, which was developed by Neungyule Education. The questions are from parts 3 and 4, as those sections have listening sections which only have answer options written in English.

The options written in English were translated into Korean in order to reproduce the same questions with Korean answer choices. For the aural English format, no words were written for the students to refer to, but test-takers could listen to an audio-recording of a native English speaker reading the possible answers. Questions with aural English options that were to be answered only from listening to a question were simply left with number options from (1)-(4) from which the students would choose the right answer, after listening to the auditory options.

The test questions were identical in meaning, except that they were presented in the three different test formats being studied. Among the 30 questions presented in each of the three tests, written Korean, written English, and aural English options were provided. Table 1 shows an example question for all 3 options, showing the method adopted for this study. As Table 1 shows, the same answer choices were required across the different formats, which all have the equivalent content.

Table 1. An Example of the Option Formats

Option types	Options provided
Written English	What type of work does the woman probably do? ① data entry ② technology sales ③ website design ④ medical research
Written Korean	What type of work does the woman probably do? ① 데이타 입력 (teyitheiplyek) ② 기술 영업 (kiswulyengep) ③ 홈페이지 디자인 (hompheyiciticain) ④ 의학 연구 (uyhakyenkwu)
Aural English	What type of work does the woman probably do? ① ② ③ ④

Qualitative analysis following the tests was completed, as the participants were requested to complete a questionnaire to gauge their perceptions of each test format (see Appendix). The questionnaire was comprised of six categories, and was written in Korean. The answers to these questions were used to address research question three ('What are the participants' preferences and attitudes towards the three different types of options?') by investigating the 169 test-takers' written responses to the survey questions. They were asked to respond using a six-point scale, with answer choices ranging from 1 (Strongly disagree) to 6 (Strongly agree). The participants were also invited to provide any general ideas and comments for each question.

3.3 Procedure

In order for the researchers to divide students into two proficiency groups, and also to prevent students of one level of English ability from being grouped together into a specific type of test format later on, the participants took a pre-test. As there were three types of test presentation formats that would be used, it was important to have all three presentation method options present within each English proficiency level. The pre-test was a listening proficiency test, which consisted of 18 multiple-choice listening questions. The students' pre-test scores are presented in Table 2 (M=6.24, SD=2.82).

Table 2. Descriptive Statistics of Pre-test Results

N	Min.	Max.	M	SD
169	1.00	16.00	6.24	2.82

In order to balance each group's English language proficiency, the participants were divided into three homogeneous groups (Group A, Group B, Group C) based on their pre-test scores. Table 3 shows the mean scores of the pre-tests of three groups.

Table 3. Mean of Pre-test Scores

Group	N	M	Min.	Max.	SD	F	p
A Written English options	59	6.31	1.00	16.00	3.13	.879	.417
B Written Korean options	59	6.20	1.00	13.00	2.72		
C Aural English options	51	6.22	3.00	16.00	2.61		

*$p<.05$, **$p < .01$*

As Table 3 shows, the mean scores of each group were very similar. Group A had a mean score of 6.31, Group B, a mean score of 6.20 and group C, a mean score of 6.22. In order to check whether the groups had an acceptable homogeneity of variance, a one-way ANOVA was run. The results indicated that there was no significant statistical difference between the three groups ($F=.879$).

Based on the pre-test, the subjects were classified into low-level and high-level proficiency groups, using the median ($Md = 6$) as the cut-off point. An independent samples t-test was used to make sure whether the groups accepted homogeneity of variance. The results in Table 4 show that the two English proficiency groups are statistically different (Low: $M=3.87$; High: $M=8.25$).

Table 4. *T*-test Results for the Pretest Scores between the Two Proficiency Level Groups

English level	N	M	SD	t
Group 1 (Low)	85	3.87	1.30	15.31**
Group 2 (High)	84	8.25	2.31	

*$p<.05$, **$p < .01$*

Two weeks after participating in this study, participants were asked to take a post-test. The participants who volunteered to take this test were given the test and the questionnaire during their regular class period during the last class of the semester. Based on the pre-test, each student was then randomly and equally distributed to take one test from among three different tests: test 1, test 2, and test 3. Group one took test 1, Group two took test 2, and Group three took test 3.

This listening test took about thirty minutes to finish. One of the three different test presentation formats being studied was given to each of the three groups. Each participant took a single test, which was made up of one of the three options. As previously mentioned, the different presentation

options given were either written Korean, written English or aural English options. Afterwards, all the participants were asked to fill out a paper-based survey directly after they completed their tests, in order to explore and understand the test-takers' opinions about the option-item formats. It took about 10 minutes for them to write responses to the questions in the survey.

4 Results

4.1 Effects of different test presentation formats on listening test performance

The results of the examination of research question one are as follows. Table 5 shows the descriptive statistics for the overall performance of the three variables: written Korean, written English, and aural English options. The aural option tests were the most difficult for the students in this experiment (*M*=3.66, *SD*=1.94). The written English option tests were in the middle (*M*=4.84, *SD*=2.10), and the written Korean option questions were the easiest (*M*=5.36, *SD*=1.91). These results indicate that there were some differences shown among the results of the three test format.

Table 5. Results of the Descriptive Statistics for Presentation of Options

Option	*N*	Min.	Max.	*M*	*SD*	Kurtosis	Skewness
Written Korean	169	.00	10.00	5.36	1.91	-.220	-.219
Written English	169	.00	10.00	4.84	2.10	-.212	.248
Aural English	169	.00	9.00	3.66	1.94	-.209	.116

To ensure whether the results were statistically significant, a one-way ANOVA with a significance level of .05 was run. The independent variable was the type of presentation option and the dependent variable was test performance. The result revealed that this test was statistically significant, as is shown in table 6 ($F = 32.246$, $p < .00$). The words in the questions in L1 most likely aided students in answering the questions.

Table 6. One-Way ANOVA Results for Presentation Options

	Sum of squares	*df*	Mean square	*F*	*p*
Between groups	255.8	2	127.9	32.246**	.000
Within groups	1999.4	504	4.0		
Total	2255.3	506			

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

Pearson's correlation coefficient was run to test whether or not there was a correlation between the type of options and the resulting scores. Table 7 shows the correlation coefficient between the options types and test scores.

Table 7. Correlation Coefficient between the Type of Options and the Scores

		Scores
	Pearson correlation	-.329[**]
Options	Sig. (2-tailed)	.000
	N	507

**. Correlation is significant at the 0.01 level (2-tailed).

As Table 7 indicates, there was a statistically significant correlation between the type of test format given and the students' scores. Thus, the answer to the first research question is that the language presentation options do have a significant impact on students' performance in listening comprehension exams.

These results are in line with those of Shohamy (1984). However, they contradict those of Rahimi (2007), where it was indicated that there was no statistical difference found in test results between items in students' target language and items in their L1. However, in contrast, Shohamy pointed out that there was a statistical difference found when giving questions in students' target language versus questions given in their L1.

As Table 8 shows, Pearson's correlation coefficient was performed again to figure out if there was a correlation between each of the three types of presentation formats and the total scores. Among the three format options, written English option scores and total scores were the most highly correlated statistically. Written Korean and aural English options were also highly correlated, but there was a lower correlation found between them than for the written English options. The correlations between written English and aural English options and written Korean and aural English options were found to be lower than the others.

Table 8. Correlation Coefficient between Three Types of Test Formats and the Scores

	Written English	Aural English	Total
Written Korean	0.425	0.400	0.780
Written English		0.355	0.784
Aural English			0.751

4.2 Listening test scores differing by test presentation format and proficiency levels

The second research question focused on whether the testing methods have a different effect according to test-takers' different proficiency levels. The participants were divided into two proficiency groups depending on their pre-test scores. An independent sample *t*-test was run in order to help analyze the second research question. The independent variable was the two different proficiency levels, and the dependent variable was the students' listening test

performance with the three presentation options. Table 9 demonstrates that the mean scores of the listening test with written Korean and written English test format were significantly different between the low-level and high-level groups. However, listening items with the aural format were not found to be statistically significant. Statistically, it can be concluded that only the written Korean and written English test presentation format options influenced the overall English listening test scores, while there was little effect found on the aural option.

Table 9. Results of *t*-test for the Effects of English Proficiency Levels on Three Different Test Formats

group	Written Korean			Written English			Aural English		
	N	M	SD	N	M	SD	N	M	SD
low	85	4.98	1.62	85	4.19	1.89	85	3.45	1.68
high	84	5.75	2.11	84	5.50	2.11	84	3.88	2.16
t		-2.67**			-4.24**			-1.46	
p		0.01			0.00			0.15	

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

As table 9 indicates, both high-level students and low-level students had the best performance on test items with written Korean options. The more advanced-level students performed better on both tests with written Korean and written English test options as compared to the aural English test option. However, lower-level students performed more successfully only on the listening tests using the written Korean test format.

In line with the present research, Shohamy (1984) pointed out that the less proficient students outperformed in items in their target language when compared with items in their L1. She pointed out that items in students' first language were more authentic to the test-takers, while items in their target language gave them more anxiety. Similar to the outcome of this research, both Rahimi (2007) and Shohamy (1984) reported that low proficiency students tended to be affected by different test formats.

4.3 Test-takers' preferences and attitudes towards the types of test formats

In order to investigate learners' preferences and attitudes towards the three different types of formats, the 169 test-takers' written responses to the survey questions were analyzed. In the survey questionnaire, the participants were asked how much they liked the written Korean option test presentation format. The questionnaire used a six-point Likert-scale, with 6 being 'like most' and 1 being 'dislike most,' with 6.0 being the maximum value and 1.0 the minimum. The survey questionnaire used in the study is shown in the appendix. The

results showed that the students liked the written Korean test format the most (M=4.92). Students chose the written English format as their second favorite choice on average (M=3.34), while aural English test options were liked the least (M=3.02). As Table 10 illustrates, regarding the question as to which test format test-takers prefer, 72% (N=121) of the participants responded that they preferred written Korean option items, 21% (N=36) preferred written English option items, and 12% (N=12) chose aural English option items.

Table 10. Test-Takers' Test Presentation Option Preferences

Options	N	%
Written Korean	121	72
Written English	36	21
Aural English	12	7

The written responses in relation to the question as to which item format test-takers preferred were revealing. To summarize their perceptions, the students who preferred the written Korean option items, many of them wrote that they did so because it was easier for them to select the correct answer. One student wrote, 'I feel less anxious with written Korean option items.' Another student wrote, 'I can read fast and save time because I don't have to translate written English options into Korean.' Furthermore, another student commented, 'They are reader-friendly.' Lastly, another participant stated, 'Before listening to the audio clip, I can guess the context.'

The participants who preferred the written English option items responded that they preferred that style of presentation for some of the following reasons: 'To improve listening ability,' 'I could find the same words written in English while I was listening,' 'I can find answers easily because some key words on the audio clip are written in English,' 'I don't have to translate Korean into English, so I can save time,' etc. Additionally, the students who preferred the aural English option format suggested such a test would better discriminate between less-able and more-able listening test-takers, or those who studied and those who did not. That is, they thought that a test using the aural English format could be, in some ways, trusted more, and better knowledge more appropriately rewarded.

With regard to learners' responses as to which of the three test formats were most appropriate and fair, and could best measure learners' listening comprehension abilities (question 3-1), table 11 indicates that aural English options (49%) were considered most appropriate to assess students' listening comprehension abilities. The written English option tests were second favorite (29%) and written Korean option questions resulted in the third highest percentage (21%). Table 11 reveals these results.

Table 11. Test-Takers' Perceptions of the Most Appropriate and Fair Item

	Written Korean	Written English	Aural English	Others	Total
N	36	49	82	2	169
%	21	29	49	1	100

Next, participants' short-answer responses regarding question 3-1 were as follows: About half of respondents mentioned that they preferred the aural English options. They preferred them for the following reasons; 'Because it is a listening test, test-takers need to choose the correct answer after listening to aural English options and aural passages,' 'To improve listening skills, test-takers should listen to options, as well,' 'To discriminate good English listeners from less skilled ones, aural English options are the fairest ones of all,' etc. On the other hand, some of them criticized aural English options, because in order to best answer items with aural English options, they needed to memorize all four options, and their concern was that the test should not be a memory test. Also, students felt that it was difficult to get helpful hints from the aural English options. The students who favored written Korean options stated that 'Aural English and written English options are too difficult to answer the questions.' The participants who preferred the written English options mentioned that 'I am used to items with written English options.' Familiarity with the written English test structure lowered some students' perceptions of the difficulty level of the test.

Concerning question 4-1 as to which test format options were the most difficult to answer, 50% (*N*=84) of the participants responded that aural English options were the most challenging. 21% (*N*=34) of them chose written English as the most difficult option. Only 7% (*N*=12) of respondents answered that the written Korean options were the most difficult. The students who chose aural English options in their questionnaire stated that written Korean options were difficult because after listening to an English script, they must then read in Korean. Therefore, they needed to translate what they heard into Korean again. They mentioned that because the listening script and questions were presented entirely in English, when the options were written in Korean, they then had two languages they needed to handle, and therefore it took them more time to translate between the two languages.

A participant who chose the written Korean options format as the most difficult mentioned that written Korean options were more challenging than she expected. The participants who chose "others" said that the type of format didn't matter because all of the options were difficult. They said that they had poor English listening skills, regardless of which test option they were presented with.

With regard to learners' responses as to which options were the easiest (question 5-1), 69% (*N*=116) of students selected written Korean options as the easiest. 12% *(N*=20) chose English options and 6% (*N*=11) chose aural

English options as the easiest type of format. Respondents who selected the written Korean options, mentioned in question 5-2 that they were used to questions with written Korean options, and they were simpler because they were in their L1. They found it easier to take the test at a faster pace. When reading Korean options quickly, they could often guess the intention of the question before they listened to the script. Of the participants who chose the English option format, one stated that written English options were easiest because they could see English words from what they heard though the recording. One participant indicated that he or she could solely concentrate on English without being distracted. Students who picked "others" (13%, N=22) pointed out that, which options they were faced with didn't matter very much, because their answers depended on their English proficiency. Many of the participants answered that all options were difficult for them.

Lastly, concerning question 6-1 as to which test format made them concentrate on the test the most during the listening test, 35% (N=59) of the participants preferred the written Korean options the most. Reasons for this included that written Korean options were the easiest, could save them the most time, while also being the most familiar to them. On the other hand, 29% (N=49) of participants selected aural English options because the test focused chiefly on listening. Since many students believed the test with aural English options was the most difficult, they felt like they needed to concentrate more in order not to miss an option and to choose the correct answer. One individual involved in the study also stated that aural English options were the most challenging because there were no written instructions or answer choices on the test. However, 21% (N=36) of all participants thought that choosing their preferred test presentation format didn't make a difference at all, because they needed to concentrate on all areas of the test. Thus, if the same question was simply being presented in a different way, the content still remained the same. Finally, 15% (N=25) of all students involved in the study ranked written English options as their preferred test method because they could answer each question at once without added translation.

4.4. Pedagogical implications

As In'nami and Koizumi (2009) pointed out, there is no ideal test which works appropriately in every setting. On that account, test developers and teachers need to acknowledge the inherent characteristics of each test format, and analyze which will best serve the purpose of each test within a certain context. An important implication of the present research is that language tests with different presentation options can influence listening test performance. The results of this study showed that there is a statistically significant correlation between the test format used and student scores, which means that scores varied depending on the types of options and how the same test questions were presented. As a result, test developers should be aware of

how each language test format may influence test scores. Therefore, for enhanced validity of test results, test makers need to consider the effects that different test methods may have, and should endeavor to reduce the influences of differing test formats in test making as much as possible. Test makers should acknowledge that test results should be fluctuated primarily based on test-takers' abilities rather than because of the test methods used.

Another pedagogical implication of this study is that students' perception of a test can differ depending on the types of test presentation formats used in an exam. The questionnaire identified that test-takers' most preferred items were items with Korean language options, mostly because they found it easier to find the right answer. In addition, among the three test formats, the items with aural English options were the least preferred items, yet they were also considered to be the most appropriate exam format to accurately assess students' listening comprehension abilities. This type of research may help English test developers and language instructors to better understand the test presentation formats they want to choose for their tests. The research presented here may help test-makers foresee the impact that making changes has on how they present question options in their exams.

5 Conclusions

A language test result is a function of two variables: test-takers' language ability and the test method used (Yi,an, 1998). The testing methods are not supposed to influence test-takers' scores. Listening test performance should solely reflect test-takers' listening abilities. However, the present research results showed that different test methods yielded different test results, which suggests that both the test method used as well as students' language abilities can have an effect on student test scores.

The results of the present study somewhat echo those of Shohamy (1984), where it was indicated that the proficiency variable affected both levels in different ways. As Shohamy (1984) reported, when it comes to written Korean and written English test options, the difference is more significant to the test methods for students who have a low proficiency level in English. An additional factor is that test questions presented in L1 reduced students' overall difficulty level. Moreover, for both written Korean and written English options, as Rahami (2007) pointed out, low English proficiency students performed more successfully when items were shown in their L1.

The study found that the mean scores were statistically significant between written Korean, written English and aural English option tests, with the written Korean option format being easiest overall. The finding that options in L1 were easier than ones using the target language may be explained as follows. First, the use of items in the first language may reduce

test-takers anxiety, which applies to both high-level and low-level groups. Second, as the qualitative analysis shows, the items presented in Korean provided students hints and clues to help them guess the correct answer. Thus, written Korean options facilitated students' understanding. Conversely, some students mentioned that answering written Korean options items took more time because they had to translate the listening script in English from English to Korean.

The present study has some limitations in that the sample size was not large enough so as to represent all test-takers. Samples might also have been truncated. Additionally, all of the students involved in the testing attended the same college and had the same majors, which means that their English performance might not vary greatly overall, as their college entrance examination scores would most likely be similar. Moreover, the limited number of multiple-choice questions given should also be considered when interpreting these results. More studies with a greater number of participants would increase the validity of this research, and more questions are needed to be posed to enhance our understanding of how language presentation formats in English listening tests affect student performance. A specific focus on which of the three formats is the most valid and reliable test method, which most accurately reflects students' English listening skills, is a topic well worth further investigation. It seems clear that additional studies are needed in this line of research, with the never-ending aim of increasing overall knowledge while enhancing the accuracy and effectiveness of language tests.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Brunfaut, T., & Revez, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly, 49*(1), 141-168.

Filipi, A. (2012). Do questions written in the target language make foreign language listening comprehension tests more difficult? *Language Testing, 29*(4), 511-532.

Hansen, C., & Jensen, C. (1994). Evaluating lecture comprehension. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 241-268). Cambridge: Cambridge University Press.

Hwang, L-S., & Lee, J-Y. (2020). Correlation analysis between the text variables and item difficulty in CSAT: Focusing on syntactic complexity. *Studies in English Language & Literature, 41*(1), 265-283.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *English Testing, 29*(2), 219-244.

Korea Institute for Curriculum and Education. (2020). Retrieved July 1, 2020, from http://www.kice.re.kr/main.do?s=english.

Lee, H., & Winke, P. (2013). The differences among the three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, *30*(1), 99-123.

Lin, Y-M., & Chen, M. Y. (2020). Understanding writing quality change: A longitudinal study of repeaters of a high-stakes standardized English proficiency test. *Language Testing*, *37*(4), 523-549.

Park, J-E., & Jang, M-J. (2019). A study on the method of providing answer options for multiple choice questions involving understanding specific information in listening and reading tests of the Korean language. *The Korean Language Literature, 144*, 331-359.

Rahimi, M. (2007). L2 reading comprehension test in the Persian context: Language of presentation as a test method facet. *The Reading Matrix, 7*(1), 151-165.

Rost, M. (2016). *Teaching and researching listening*. New York: Routledge.

Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, *37*(1), 31-53.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *2*(1), 199-215.

Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalizing the test construct. *Journal of English for Academic Purposes, 10*(2), 89-101.

Yi'an, W. (1998) What do tests listening comprehension test?: A retrospection study of EFL test takers performing a multiple-choice task. *Language Testing, 15*(1), 21-44.

Appendix
Questionnaire

This questionnaire will be used only in an English listening test study. Thank you for your participation.

1. (A-C) Circle your answers depending on your preference.

strongly disagree   ←   →   strongly agree
            1   2   3   4   5   6

A. I prefer written Korean options.	1	2	3	4	5	6
B. I prefer written English options.	1	2	3	4	5	6
C. I prefer aural English options.	1	2	3	4	5	6

2-1. Which one do you prefer most? Why do you prefer?

① Written Korean options ② Written English options

③ Aural English options   ④ Others

2-2. Explain why you chose your answer.

_____

3-1. Which options are most appropriate and fair to measure learners' listening comprehension abilities?

① Written Korean options ② Written English options

③ Aural English options   ④ Others

3-2. Explain why you chose your answer.

_____

4-1. Which options are most difficult?

① Written Korean options ② Written English options

③ Aural English options   ④ Others

4-2. Explain why you chose your answer.

_____

5-1. Which options are easiest?

① Written Korean options ② Written English options

③ Aural English options   ④ Others

5-2. Explain why you chose your answer.

_____

6-1. Which options made you concentrate on the test the most during the listening test?

① Written Korean options ② Written English options

③ Aural English options   ④ Others

6-2. Explain why you chose your answer.

_____

Jinseon Koo, Ph.D. Candidate
Department of English Education
Ewha Womans University
Ewhayedae-gil 52, Seodaemun-gu, Seoul, Korea
E-mail: jinseonkoo@ewhain.net

Eun-Joo Lee, Professor
Department of English Education

The Effects of Differing Language Test Presentation Methods on
EFL College Students'Listening Test Performance

Ewha Womans University
Ewhayedae-gil 52, Seodaemun-gu, Seoul, Korea
E-mail: eunlee@ewha.ac.kr