A Study of High School English Listening Assessment Items Based on a Discriminant Analysis*

Nayu Kim · Heechul Lee*

(Jeonbuk National University)

Kim, Nayu & Lee, Heechul. (2020). A study of high school English listening assessment items based on a discriminant analysis. *The Linguistic Association of Korea Journal, 28*(2), 1–14. The purpose of this study is to investigate what types of listening assessment items take an important role in identifying high school student's English listening ability, and to explore teacher's considerations about designing and modifying the types of items. 187 high school students participated in this research and took an English listening test made up of twenty multiple-choice items. Each student's score was unpacked by discriminant analysis, and it was investigated what types of items made a difference when the students were classified into three English listening levels; high, intermediate or low. Two types that had a key role on distinguishing high level listeners from the others were "to identify the other person's response" and "to infer what one asks the other." In contrast, low level listeners responded appropriately to such types as "to find out main idea of a dialogue" and "to guess character's following response after a dialogue." Based on the findings from this research, the English teacher can identify a high or low level listener easily and correctly based on the above four types of listening test items. In addition, the English teacher should ensure the validity and reliability of the items, and modify them properly considering a student's cognitive and affective development, listening strategies and learning motivation in order to construct the effective listening assessment system.

Key Words: English listening assessment item, English listening level, discriminant analysis

1. Introduction

English listening is a substantial component of English communication skills (Brown, 2005). When a linguist or an English teacher explains four skills including listening, speaking, reading and writing, it is not uncommon for him or her to prefer listening first (Dekeyser, 2007). Any chapter of an English textbook or an ESL (English as a Second Language) reference normally begins from a listening section, and then speaking and reading parts follow that (Oxford, 1990). It is not just on a composition of a book. Listening assessment, as reading one, has a significant role of any English placement, formative or summative test (Kim, 2014). For example, among forty five items in CSAT (College Scholastic Aptitude Test) English portion in Korea, seventeen items are directly related to listening and speaking skills, which have the format of one or two native speakers' utterance and student's answer for the question from that, and are largely acknowledged as listening test due to their pattern.

There are some limitations to evaluating a student's listening ability systematically, even if a student exposes to circumstances that he or she focuses on listening on the preferential basis in the English curriculum (Graham, 2006). All of English teachers as non-native speakers in ESL or EFL (English as a Foreign Language) course necessarily prepare for audio clips like MP3 audio files recorded for listening comprehension if they do not work with a native English teacher. It is not always easy for them to design an effective listening formative assessment

---

** The first author is Nayu Kim and the corresponding author, Heechul Lee.

in the context where the audio files cannot be modified freely due to technological limitations like editing or re-recording them. Naturally, an English teacher prefers already made listening comprehension exercises to producing new items, and it cannot be that bad in respect to their universal validity, reliability and efficiency (Jung, 2010). Considering this context, listening test items have an important role, in that they are very close to both a teacher and a student and are a direct and frequent measurement to assess a student's listening ability. There have been a lot of types of listening item that many linguists and English teachers have developed based on various English learning and teaching principles (Yo & Lie, 2009). In addition, the teacher should reflect various factors such as a creative teaching approach and a meaningful assessment strategy when designing the English listening items (Anna & Shyang, 2008). In the same veins, as every type of items has distinct merits and demerits, an English teacher can use one on various purposes. On the other hand, an English teacher who should figure out his or her students' listening ability in a limited period time is supposed to be confused with deciding what types of items are effective (Alderson, 2000). In order to identify the key item of the total, it is necessary to analyze individual items in a scientific research method and to investigate statistically what item is effective on assessing a student's listening ability.

It is efficiently identifying a high or low level listener that a most significant role that a listening test item has is (Field, 2013). If an English teacher knows effective types of item to assess students' listening ability easily and fast, it is useful for him or her to design a lesson plan fit for them. In addition, an English teacher can get a clue where to watch when he or her want to manufacture a proper assessment item. To obtain these goals in this study, the following research questions are central to a discriminant analysis: 1) what is the most important type of items in distinguishing a high or low level listener from others? and 2) what should an English teacher keep in mind when he or she designs that type of items?

2. Theoretical Background

2.1 Approach of English Listening Assessment

Katz (2012) suggests that teachers should make inferences about the development of listening skills since listening processing is not an observable behavior in the same way as speaking or writing a language. In particular, listening assessment are ubiquitous and can be found anywhere, from elementary schools to high schools, colleges and universities, the workplace, and even immigration and naturalization context, and at all levels, form beginning to advanced levels of language teaching and learning. In these different contexts, the language assessments used could be large-scale assessments (such as those prepared by professionals in testing agencies, a university, or a state board or ministry of education) or small-scale assessments (such as those prepared by a teacher, an assessment development group or committee, or a single immigration or naturalization examiner). Teachers in various educational conditions need to understand the characteristic, effect and function of each test item (Kunnan, 2009). In the same vein, Hughes (2003) argues that the challenge for the assessment of receptive skills like listening is to set tasks which will not only cause the candidate to exercise listening skills, but will also result in behavior that will demonstrate the successful use of those skills.

The discussion of the objects of listening assessment has been carried out on the various dimensions. It is a substantial issue on the validity of listening assessment whether only aural information should be included in listening assessment procedure or visual description could be presented with aural information in it. Ockey (2007) suggests that the decision of whether the developers and the users contain visual materials in listening assessment should be about what construct investigated by the assessment is. Some researchers argue that listening ability should be evaluated by just aural voice since listening is a kind of aural phenomenon (Buck, 2001). Meanwhile, others want to expand the definition of listening assessment, which

includes not only aural information but also visual one carried by speakers (Field, 2013; Gruba, 1997). For example, Field (2013) insists that the visual information of spoken language is not for just carrying additional and complementary intelligence but an inevitable component in listening comprehension. The test developer in the course of designing listening evaluation system should define "listening competence" as a construct of test, when an inevitable consideration is whether nonlinguistic components, like visual thing, are included in the assessment construct (Batty, 2015; Buck, 2001; Cubilo, 2017). In addition, Carr (2011) suggests that more components like the pattern on utterance, the communicative function, the use of super-differentiated sound, the accent, and language register need to be considered when discussing listening assessment even though the genre and rhetoric structure is important in reading assessment. In the same vein, Park (2019) points out that the type of the spoken text can be classified into various dimensions by the degree how it is presented with the nonlinguistic components, which is a very important factor when considering using what pattern of spoken language resource in the development of listening test.

In the field of listening assessment methods, Alderson (2000) points out that no single test method can achieve all purposes for which we may assess in language assessment. As the methods of English listening assessment is various depending on its examinee, the endeavor to analyze listening assessment formats has been continued broadly in a global scale. For example, Lee, Lee and Jun (2016) investigate TOEFL (Test of English as a Foreign Language), TOEIC (Test of English for International Communication), IELTS (International English Language Testing System) and PTE (Pearson Test of English) Academic familiar with Korean examinee based on their purpose with subject. According to their research, TOEFL has the format for evaluating English ability of those who do not use English as their mother tongue, the purpose of which is to assess English ability of international students to hope to study in the college in North America. TOEIC is the English test developed to check English communication ability of people who work in business setting, the result of which reveals how well the examinee communicate with the partner in the international working group. IELTS is the English assessment for people who want to go to study, immigrate or be employed in English speaking countries. PTE(Pearson Test of English) Academic is the test for a person's English ability in academic setting, similar to TOEFL and IELTS in its purpose (Lee, Lee and Jun, 2016).

2.2 Study of English Listening Test Item

The research of the validity of English listening test for adult, for the CSAT English portion based on the National Curriculum and for English achievement test for secondary school students which was designed by the local Ministry of Education has been performed on an item-focused basis (Kim, 2007; Kim, 2006; Jung, 2006). In addition, Kim (2006) investigates the quality like validity and reliability of English listening test items made by the University devoted to foreign language education in Korea, where the level of item difficulty, item discrimination power and guessing parameter of item are compared based on IRT (Item Response Theory). Meanwhile, Han and Lee (2010) analyze the validity of a nationwide English listening achievement test item on high school students, who argue that the percentage of correct answers is closely related to the obstructive factors of listening like speed of speech, the sound system, the complexity of syntactic structure, the familiarity of meaning of words and the placement of information. Jang and Kim (2016) investigate the effect of the nationwide listening test items for the middle school students. They insist that it is possible for an English teacher to perform the measurement-driven listening instruction by linking classroom activity into listening test.

Of various English listening assessment items, the items of the CSAT English portion have received attention from researchers since the test has an important part in the college admission procedure in Korea. Jung (2010) reviews the validity of the listening test items of the CSAT English portion conducted from 1994 to 2010, where all items are classified into five domains including the language function, the communication ability, the topics used in the item, the type of language materials and the pattern of

evaluation. In that research, it is revealed that the listening items is closely related to the National English Curriculum but it is necessary for the item to be set in a balance way in respects to topics and types. Chang (2012) conducts the research of the construct validity of the listening test items of the CSAT English portion and of the local Ministry of Education, where the confirmatory factor analysis is used to investigate the validity and reliability. He suggests that listening test items need to be modified to develop the student's understanding and adaptability. Lee (2001) classified the listening test items of the CSAT English portion into various types based on the way that the items focused on factual understanding about whether the examinee finds out the specific information in the script of item such as time, place, fare, job, major and purpose. In addition, he claims that it is necessary to be careful for the effect of complexity on item difficulty when developing the type of items since some items ask the student to have the competence to infer in a long script.

In the dimension of detailed item analysis or response, Brown and Abeywickrama (2010) categorize listening test items into various types based on the topic that is addressed heavily in that question. For example, they suggest that the type of identifying time and number presented from the script is divided into the type of simple calculation by short memory and the type of complicated calculation by enough insight. Haladyna (2016) insists on the importance of the analysis of options in the listening item since the quality of the item becomes good when the options of the item are made properly, and suggests that an examiner should be given the opportunity that replace and modify the options of the item since ineffective options don't have an effect that the developer intends. In addition, she argues that item analysis needs to be conducted in order to identify whether the response to individual listening items is valid or not.

3. Research Method

3.1 Participants

One hundred eighty seven high school students participated in this research. All of them were male and 12th graders in the same high school and were 18 or 19 years old. They are divided into six classes when they had official English ones. Their English achievement test grades were in normal to level 4 or 5 of nine levels, which had been evaluated based on official English assessment instruments by the Office of Education. They had five English classes as an official curriculum and one or two English ones of extracurricular classes a week. There were no students who had been educated in foreign countries for more than six months. In this study, top 21.3% students were regarded as high level (Level 1) based on their grades of English listening test using the instrument in this research as shown in Table 1 because the cumulative percentages of three high levels of total nine levels in the CSAT English portion are 23%.

Table 1. Participants' English Listening Level

English listening level		Students	Percent (%)	Cumulative (%)
High	Level 1	40	21.3	21.3
↕	Level 2	62	33.1	54.4
Low	Level 3	85	45.6	100
	Total	187	100	

In the same vein, the lower 50% or so of the total students were set to the lower level (Level 3), and the upper 23% to 50% were set to the middle level (Level 2). Students of level 1 had more than 16 points of total 20 ones and were 40 students, 21.3% of 187 students. Level 2 meant the students who had more than 11 to 15 points and ones who got under 10 points were assigned to level 3. Sixty two students were in Level 2, 33.1% and eighty five students in Level 3, 45.6%.

3.2 Instrument

The instrument consisted of twenty multiple-choice items that were classified into different types and language of options, Korean or English. An individual item was designed by Education Broadcasting System, a public organization of education, and many high school students took the test in order to confirm the listening ability by themselves. All items had a script to handle a communicative situation or a dialogue, and item 15 and 16 had the same one. Each statement of an item took about one minute, so total test time was twenty minutes or so. The options including a right answer and four distractors of total items were in Korean or English, eight items of which were written in Korean and the other ones in English.

The types of listening items were classified based on various criteria for each researcher. Lee (2001) categorized CSAT listening items by the patterns and contents in the scripts. First, the types that measured factual understanding were the items that asked for a clear understanding of the facts of what is being heard, such as time, place, rate, job, major, and purpose, or details. Second, the types evaluating inference ability were the items of identifying the content that was not explicitly indicated in the scripts. The types included "to request", "to identify contents and topics", "to identify relationships and occupations", "to identify conversation places", and "to find purposes and intentions." Third, the types that measured overall understanding consisted of the items that judged the mood, reconstructed information, or asked emotions such as "to choose the right response for the situation" and "to guess the feeling." Fourth, the types of indirectly evaluating speaking skills included the items of "to complete conversation", "to convert discourse into conversation", "to summarize conversation content", and "to understand and respond to conversation situations." The English listening items used in this study were classified based on the Lee (2001)'s criteria of the CSAT item types, and for items with ambiguous classification, it was modified so that the characteristics of the items could be expressed correctly.

Table 2. Item Types

Item	Item Type	Language of Options
Item 1	to identify the other person's response	English
Item 2	to identify the other person's response	English
Item 3	to identify the other person's response	English
Item 4	to understand objective of a speech	Korean
Item 5	to understand the speaker's point of view	Korean
Item 6	to find out a main idea of a monologue	Korean
Item 7	to find out a main idea of a dialogue	Korean
Item 8	to guess the place where a dialogue takes place	English
Item 9	to identify a wrong statement on bar graphs	English
Item 10	to guess the following activity after a dialogue	Korean
Item 11	to infer what one asks the other	English
Item 12	to find out a specific information of a dialogue	Korean
Item 13	to find out a wrong statement of a dialogue	Korean
Item 14	to calculate total cost of merchandise purchased	English
Item 15	(two items to understand the speaker's point of view	English
Item 16	on a speech) to identify a wrong statement on a speech	Korean
Item 17	to find out a right information on a table	English
Item 18	to identify the other person's response	English
Item 19	to identify the other person's response	English
Item 20	to guess character's following response after a dialogue	English

3.3 Data Analysis

The participants' test grade for the instrument was entered into the SPSS 20.0, a statistical program, for a multivariate analysis. The discriminant analysis used in this research is the appropriate statistical technique when the dependent variable is a categorical (nominal or nonmetric) variable and the independent variable was metric variable. Discriminant analysis involves deriving a variate. The discriminant variate is the linear combination of the independent variables that distinguish best between the objects in the groups (Hair et al., 2010). In this study, when a student chose a right answer for an item, the result was coded as

"1" on the item, and "0" for a wrong answer, which was a nominal variable as the independent one, and the dependent variable was level 1, 2 or 3 as a metric variable that an individual student was put together. In other words, it was the key point that what types of items affected on distinguishing a level group from the other ones. In addition, what types of item took an important role in an individual group was analyzed by discriminant functions extracted. The effect of each item in an individual level was investigated by various discriminant functions in this study too. In discriminant analysis, more than one discriminant function may be present, resulting in each object possibly having more than one discriminant score.

4. Results and Discussion

4.1 Response to An Individual Item and Test of Box's M

The students of level 1 who had high listening ability guessed a correct answer on most of the items. In particular, all of them found a right answer on item 1, 2, 3, 5 and 18, which were "to identify the other person's response", "to understand the speaker's point of view" and "to identify the other person's response" as shown in Table 3. However, there were three items that they had a difficulty dealing with, which were item 8 "to guess the place where a dialogue takes place", item 9 "to identify a wrong statement on bar graphs", and item 15 "to understand the speaker's point of view", and the items had less than .60 of mean value. Item 15 that was the most difficult to level 1 was the similar type to item 18 that was analyzed as an easy one, but they had clearly different pattern in that item 15 and 16 were included in the same script as a group. Likewise, the students of level 2 reported good grades, more than .9 of mean value, on item 1, 2, and 18 and had a trouble finding out the right answer on item 8, 9 and 15 too. The mean value of item 15 in level 2 was .08, which was the lowest of total items regardless of level. In level 3, the students took low grade on whole items, and had strength and weakness in the items similar to level 1 and 2.

Table 3. Mean and S.D of an Individual Level on Items

Item	Level 1		Level 2		Level 3	
	Mean	S.D	Mean	S.D	Mean	S.D
Item 1	1	0	.94	.24	.67	.47
Item 2	1	0	.92	.27	.78	.41
Item 3	1	0	.61	.49	.34	.47
Item 4	.88	.33	.61	.49	.67	.47
Item 5	1	0	.77	.42	.55	.50
Item 6	.98	.15	.77	.42	.56	.49
Item 7	.98	.15	.89	.31	.65	.48
Item 8	.60	.49	.37	.48	.13	.33
Item 9	.60	.49	.15	.35	.12	.32
Item 10	.85	.36	.58	.49	.20	.40
Item 11	.95	.22	.56	.50	.28	.45
Item 12	.85	.36	.58	.49	.27	.44
Item 13	.95	.22	.68	.47	.47	.50
Item 14	.88	.33	.65	.48	.40	.49
Item 15	.43	.50	.08	.27	.14	.35
Item 16	.88	.33	.66	.47	.39	.49
Item 17	.80	.40	.56	.50	.26	.44
Item 18	1	0	.94	.24	.64	.48
Item 19	.98	.15	.77	.42	.18	.38
Item 20	.93	.26	.84	.37	.51	.50

The Box's M is a statistical test for the equality of the covariance matrices of the independent variables across the groups of the dependent variable. If the statistical significance does not exceed the critical level, then. The equality of the covariance matrices is supported. If the test shows statistical significance, then the groups are deemed different and the assumption is violated (Hair et al., 2010). The statistical significance in this research was .028, which did not exceed the critical level, and the discriminant analysis

was regarded as an acceptable one as shown Table 4.

Table 4. Test of Box's M

Box's M		295.209
F	Approx.	1.195
	df1	210
	df2	53076.982
	Sig.	.028

4.2 Discriminant Functions

Discriminant function consists of an intercept, discriminant coefficient and independent variables as the form of the linear equation (Hair et al., 2010). Simultaneous estimation was chosen in this research, which involves extracted the discriminant function so that all independent variables are considered simultaneously. Thus, the discriminant function was computed based upon the entire set of independent variables, regardless of the discriminating power of each independent variables (Hair et al., 2010). The simultaneous estimation was selected in this study was because it was necessary to include all items as the independent variable and not to be interested in seeing intermediate results based only on the most discriminating variables (Hair et al., 2010). The eigenvalue of this analysis was 6.20, the variance was 95.8% and the canonical correlation was .92 as shown in Table 5, which meant discriminant function 1 explained 95.8% of whole variance in this analysis and the model fit was acceptable.

Table 5. Eigenvalue

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.20	95.8	95.8	.92
2	.27	4.2	100.0	.46

The statistical significance of the test of functions in this research was .001 or .002 as shown in Table 6, which was considerably good ($p \leq .05$). The Wilks' Lambda is many times referred to as the multivariate F and in commonly used for testing overall significance between groups in a multivariate situation. It considers all the discriminant functions; that is, it examines whether groups are somehow different without being concerned with whether they differ on at least one linear combination of the dependent variables. If the Wilks' Lambda is "1" of the independent variables, the mean value of the dependent variables is same. In other words, the lower the Wilks' Lambda is, the higher the difference between the mean values of the dependent variables is, which means the low Wilks' Lambda is good for prediction of cases (Hair et al., 2010). The Wilk's Lamda on Function 1 was .11, which was statistically significant because it was considerably low.

Table 6. Test of Functions

Function	Wilks' Lambda	Chi-square	df	Sig.
1	.11	386.43	40	.001
2	.78	41.81	19	.002

The discriminant coefficient is a weight whose size relates to the discriminatory power of that independent variable across the groups of the dependent variable. Independent variables with large discriminatory power usually have large weights, and whose with little discriminatory power usually have small weight. The relative magnitude of one variable if added to the other variables already in the equation can be assessed by the standardized coefficient (Hair et al., 2010). What had high coefficient in Standardized Function 1, higher than .3, was item 3 "to identify the other person's response", item 5 "to understand the speaker's point of view", item 10 "to guess the following activity after a dialogue", item 11 "to infer what one asks the other" and item 19 "to guess character's following response after a dialogue" as

shown in Table 7, which meant five items comparably took an important part in distinguishing a level from the other ones. In particular, item 11 and 19 were just two variables higher than .4, which implied that students had a difficulty inferring a proper response in dialogue based on the context. This type of items asked the student to process the information from a dialogue based on the context, not to memorize whole specifications. Otherwise, item 7 "to find out a main idea" and item 9 "to identify a wrong statement on bar graphs" did not effect on that, the type of which featured in repetition of key words to present a clue in the script and the student's sustainable check on the illustration even if the statement of the script passed by.

Table 7. Canonical Discriminant Coefficients

Item	Unstandardized Function 1	2	Standardized Function 1	2
Item 1	.39	−.32	.13	−.11
Item 2	.65	−.37	.21	−.12
Item 3	.92	.34	.39	.14
Item 4	.31	.73	.14	.33
Item 5	.76	.04	.31	.01
Item 6	.93	.02	.39	.00
Item 7	.31	−.16	.11	−.06
Item 8	.59	.26	.25	.11
Item 9	.36	1.31	.13	.49
Item 10	.80	−.27	.34	−.11
Item 11	.96	.63	.41	.27
Item 12	.64	.12	.28	.05
Item 13	.42	−.08	.18	−.03
Item 14	.51	.03	.23	.01
Item 15	.65	1.34	.24	.49
Item 16	.63	.03	.28	.01
Item 17	.52	−.05	.23	−.02
Item 18	.23	−.54	.08	−.19
Item 19	1.23	−1.11	.44	−.40
Item 20	.49	−.39	.20	−.16
Constant	−7.10	.44		

The discriminant loading is a measurement of the simple linear correlation between each in independent variable and the discriminant Z score for each discriminant function, which is calculated whether or not an independent variable is included in the discriminant function (Hair et al., 2010). Item 3, 10, 11, and 19 that had higher standardized coefficient had significant correlations in the Function 1 as shown in Table 8. The items had two things in common, which were the script of item was based on a dialogue, not monologue or one's presentation, and asked them indirect contextual information, not a concrete specification.

Table 8. Structure Matrix

Item	Function	
	1	2
Item 10	.24	−.08
Item 11	.23	.13
Item 3	.23	.15
Item 12	.20	−.01
Item 17	.19	−.04
Item 8	.17	.01
Item 16	.17	−.03
Item 5	.17	.02
Item 13	.16	.08
Item 14	.16	.00
Item 6	.15	.01
Item 2	.11	−.07
Item 9	.17	.52
Item 15	.09	.50
Item 19	.37	−.43
Item 4	.05	.32
Item 18	.17	−.27
Item 20	.17	−.23
Item 1	.16	−.23
Item 7	.14	−.15

The centroid is mean value for the discriminant Z scores of all objects within a particular category or group. For example, a two-group discriminant analysis has two centroids, one for the objects in each of the two groups (Hair et al., 2010). Three centroids of three level groups were extracted, of which the centroid of level 1 was the highest, 3.85 and one of level 3 was the lowest, −2.43 as shown in Table 9. The scatter plot based on the discriminant functions and the centroids was presented as Figure 1, where all of students gathered clearly in three groups. In particular, the group of level 1 was most concentrated, with little overlap with the other two groups, which meant the discriminant analysis was statistical significant in distinguishing high level listener from others.

Table 9. Group Centroids

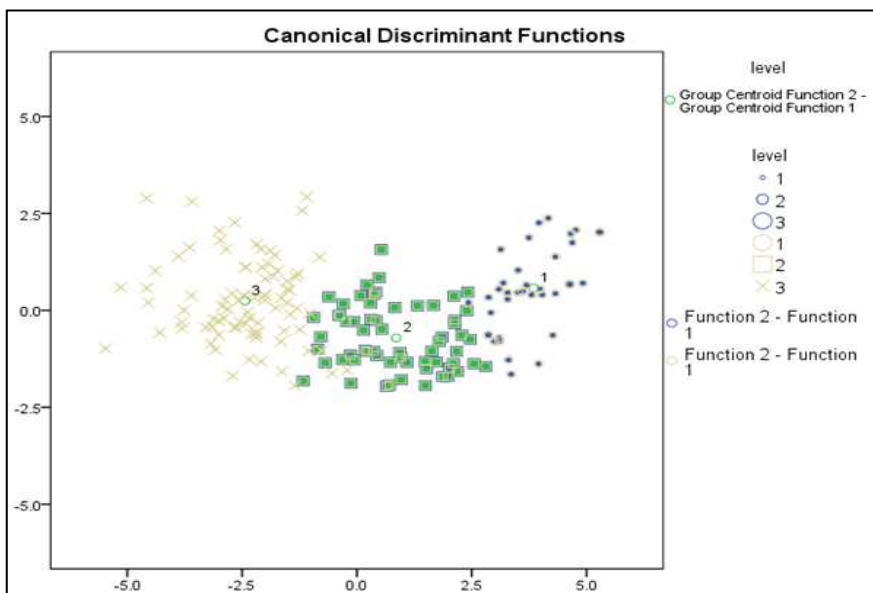| Level	Function 1	Function 2
1	3.85	.57
2	.85	−.71
3	−2.43	.24

Figure 1. Scatter Plot of Canonical Discriminant Functions

4.2 Classification of Discriminant Analysis

Classification function is a method of classification in which a linear function is defined for each group. Classification is performed by calculating a score for each observation on each group's classification function and then assigning the observation to the group with the highest score. It differs from the calculation of the discriminant Z score, which is calculated for each discriminant function (Hair et al., 2010). Item 2 "to identify the other person's response", item 6 "to find out a main idea of a monologue", and item 11 " to infer what one asks the other" were the most important in the classification function of level 1 as shown in Table 10. The item types were focused on a student's power of deduction from context, not a simple information. The student could identify a correct answer considering the logic of the statement, contextual situation, whole information even though he or she missed a word or phrase at that time. However, it was only possible when he or she had a listening strategy and whole communication ability in English that could not be learned for a short time. The function of level 1 including discriminant coefficients and a constant was as follows:

$$\text{Function of Level 1} = -66.76 + 6.64 \times \text{item 1} + 12.57 \times \text{item 2} + 9.9 \times \text{item 3} + 5.89 \times \text{item 4} + 9.46 \times \text{item 5} \ldots\ldots + 7.78 \times \text{item 20}.$$

In addition, item 2 "to identify the other person's response", item 5 "to understand the speaker's point of view" and item 6 "to find out a main idea of a monologue" very affected on the classification function of level 2. Item 5 was the type similar to item 6, which was focused on inferring speaker's intention as a main idea. Item 5 and 6 were a little bit easy in that the language of options on each item was Korean, which meant the students of level 2 got the right answer consistently based on their listening ability regardless of understanding the options. The function of level 2 including discriminant coefficients and a constant was as follows:

$$\text{Function of Level 2} = -39.07 + 5.87 \times \text{item 1} + 11.11 \times \text{item 2} + 6.67 \times \text{item 3} + 4.01 \times \text{item 4} + 7.12 \times \text{item 5} \ldots\ldots + 6.82 \times \text{item 20}.$$

Item 2 "to identify the other person's response" had a significant role in the classification function of level 3 and its coefficient was 8.61, which had considerable difference in the other items. This kind of item was similar to item 1, 3, 18, 19, 20 in that all of them asked the students to guess the response for a statement, but the length of the script of item 2 was comparably short and the topic was familiar with students' daily life, so that the listening ability of the students of level 3 was identified clearly. The function of level 3 including discriminant coefficients and a constant was as follows:

$$\text{Function of Level 3} = -17.65 + 4.27 \times \text{item 1} + 8.61 \times \text{item 2} + 3.96 \times \text{item 3} + 3.7 \times \text{item 4} + 4.64 \times \text{item 5} \ldots\ldots + 4.82 \times \text{item 20}.$$

Table 10. Classification Function Coefficients

Item	Level 1	Level 2	Level 3
Item 1	6.64	5.87	4.27
Item 2	12.57	11.11	8.61
Item 3	9.90	6.67	3.96
Item 4	5.89	4.01	3.70
Item 5	9.46	7.12	4.64
Item 6	10.24	7.40	4.33
Item 7	6.78	6.05	4.86
Item 8	5.71	3.57	1.86
Item 9	3.42	.63	.68
Item 10	8.13	6.07	3.16
Item 11	10.18	6.47	3.91
Item 12	5.96	3.87	1.87
Item 13	4.58	3.42	1.96
Item 14	5.93	4.35	2.68
Item 15	7.97	4.28	3.41
Item 16	6.36	4.42	2.38
Item 17	5.33	3.83	2.06
Item 18	4.93	4.91	3.61
Item 19	9.15	6.87	1.74
Item 20	7.78	6.82	4.82
Constant	−66.76	−39.07	−17.65

The results of the classification procedure were presented in matrix form, as shown in Table 11. The entries on the diagonal of the matrix represented the incorrect classifications. The entries under the column labeled "Original Group" represented the number of individuals actually in each of the three groups. The entries at the bottom of the columns represented the number of individuals assigned to the groups by the discriminant function and the percentage correctly classified for each group was presented too (Hair et al., 2010). The discriminant function correctly predicted 82 students of 85 ones of level 3, 96.5% and 59 students of 62 ones of level 2, 95.2% and 40 students of 40 ones of level 1, 100%, which meant 96.8% of original grouped cases correctly classified.

Table 11. Classification Results

	Original Group	Predicted group membership			Total
		Level 1	Level 2	Level 3	
Count	Level 1	40	0	0	40
	Level 2	2	59	1	62
	Level 3	0	3	82	85
	Total	42	62	83	187
Percent (%)	Level 1	100	0	0	100
	Level 2	3.2	95.2	1.6	100
	Level 3	0	3.5	96.5	100

5. Conclusion and Implication

In this research, the discriminant analysis was performed in order to explore the effect of the type of English listening assessment items on distinguishing a high, intermediate or low level listener from others. The discriminant functions were extracted based on the simultaneous estimation, and the items that had an important role in a teacher figuring out a student's listening level were investigated by the discriminant coefficients and the discriminant loadings of the functions. In addition, the individual discriminant functions of three groups of levels were analyzed and the specific items closely related to each level were identified too. Considering the results of this research, it is possible to identify the characteristics of critical items to effect on each level listener, and the findings suggest implications of designing and modifying that type of items to an English teacher as below.

The items that had higher discriminant coefficient and discriminant loading were item 3 "to identify the other person's response", item 10 "to guess the following activity after a dialogue", item 11 "to infer what one asks the other" and item 19 "to identify the other person's response." In level 1, the high level

listener's group, item 2 "to identify the other person's response", item 6 "to find out a main idea of a monologue" and item 11 had a significant role. In addition, in level 2, the intermediate level listener's group, item 5 "to understand the speaker's point of view was important with item 2 and item 6, and in level 3, the low level listener's group, item 2 was the most statistically significant in identifying the student.

As a result of analyzing each discriminant function, the type of "to identify the other person's response" was the most important on all three levels, which featured in a comparably short dialogue in English, a familiar topic closely related to a student's daily life and a clear expression of agreement or disagreement. Therefore, it is necessary for an English teacher to thoroughly check the logic of the script and to selectively use the vocabulary having clear meaning since the statement in the type of item passed by so fast and a complex or ambiguous word misled the context in an unpredictable way. In particular, the distractors of the options are better to clearly express consent or not since just a word placed first in a response affects considerably on the student understanding the following clause.

In the group of high level listeners, called level 1, the types of "to find out a main idea of a monologue" and "to infer what one asks the other" were investigated as distinctive features by the discriminant function. Two types have characteristics in common that ask the student to sum up all the information and to identify a logical conclusion as a main idea or a following statement. Memorizing specific information in the script is not necessary to guess the right answer, and it is important to keep consistency and cohesion of the entire statement in that type. Hence, a teacher had better focus on the authentic composition of the script when designing the types of item. It is not effective to include excessively difficult expressions in the item in order to figure out who the high level listener is. In addition, a teacher needs to be careful when he or she chooses a topic on it. However the contents in the item are logical, a student is not capable of absorbing the situation if the script is closely related to an unfamiliar topic.

Meanwhile, in the group of low level listeners, level 3 and item 5 "to understand the speaker's point of view" had an important role with item 6 the same as level 1. This type has a correct answer and four distractors in Korean, which minimize the difficulty from a student's processing English options. The student keeps attention to audio information and tries to find out repetitive phases on the script as key words directly related to the main idea. When a teacher makes this type of item, it is important to design and modify carefully the Korean options. Since the difference of semantic features between English and Korean, a teacher needs to try to avoid ambiguity and confusion from two languages in one item. In addition, the speaker's point of view is presented like a claim, which means the topic needs to be familiar with the student and to be associated with something positive for educational aims.

Summarizing the findings discussed above, it is relatively clear that students with different English listening levels respond individually to the specific types of items. Just as there are a variety of factors that affect students' English listening, the response of items based on listening level also shows distinctive characteristics. In order for teachers to effectively present listening items to students and use the results, it is necessary to develop and revise various types of listening questions. In-depth analysis of textbook-based listening items and CAST items that are currently used in most English classes needs to be performed, and based on this, creative and effective items that match the reality of English class in Korea are needed. In addition, teachers should design and apply listening test items optimized for each individual student in consideration of cognitive and emotional states as well as students' English abilities when producing listening items. As mentioned above, the degree of response to the items varies depending on the level of students' English listening, which means that students' responses may vary by more diverse factors. Therefore, teachers need to design and utilize English listening items that appropriately reflect the characteristics of students while continuously monitoring students' motivation to participate in English classes, learning types, and learning strategies.

The participants in this research were only high school students, all of whom were male. It is necessary to conduct expansive research whose subjects are in elementary school, junior high school and college, or

female students in order to analyze deeply the feature of listening assessment items. In addition, the quantitative research methods like cluster analysis, multiple regression and structural equation model need to be used on various aspects of exploring individual items since there are many contextual components in approaching listening assessment. Likewise, an English teacher's acknowledgement for designing listening test items need to be investigated using qualitative study.

References

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

Anna, C., & Shyang C. (2008). Listening strategies of L2 learners with varied test tasks. *TESL Canada Journal 25*(2), 15-16.

Batty, A. O. (2015). A comparison of video-and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing, 32*(1), 3-20.

Brown, D. H., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd Eds.). NY: Pearson Education.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*, New York: McGraw-Hill College.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University Press.

Chang, J. (2012). The construct validity of models of L2 listening ability. *Modern English Education, 13*(2), 207-229.

Cubilo, J. (2017). *Video-mediated listening passages and typed note-taking: Examining their effects on examinee listening test performance and item characteristics* (Unpublished doctoral dissertation). University of Hawaii at Manoa, Honolulu, HI.

Dekeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: CUP.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77-151). Cambridge: Cambridge University Press.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System, 34*, 165-82.

Gruba, P. (1997). The role of video media in listening assessment. *System, 25*(3), 335-345.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New Jersey: Pearson Education.

Han, M. S., & Lee, J. S. (2010). An analysis of factors hindering English listening comprehension: Based on the English listening comprehension test of high school students. *Secondary English Education, 3*(1), 107-135.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.

Jang, S. Y., & Kim, T. (2016). Exploring Validity of the Nationwide Standardized English Listening Tests Using the Rasch Measurement Model. *Studies in English Education, 21*(2), 115-145.

Jung, H. (2010). An analysis of spoken English test items in Korean CSAT. *Studies in English Education, 15*(2), 177-197.

Katz A. (2012). Linking assessment with instructional aims and learning. In C. Coobe, P. Davidson, S. Stoynoff, & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research on language education*. Cambridge, UK: Cambridge University Press.

Kim, D. J. (2014). Analysis on learning effects of a university English class partially-integrating reading and listening subjects. *Modern Studies in English Language & Literature, 58*(2), 33-65.

Kim, G, S. (2007). A study of a teaching method for high score strategy of listening assessment of high school

English learners. *STEM Journal, 8*(1), 141-168.

Kim, I. (2006). Effects and agenda of the nationwide English listening comprehension test in Korea. *The Linguistic Association of Korea Journal, 14*(1), 183-202.

Kunnan, A. J. (2009). The U.S. naturalization test. *Language Assessment Quarterly, 6*, 89-97.

Lee, B., Lee, Y. J., & Jun, H. (2016). The validity of the new item types of the listening section of an English proficiency test. *Secondary English Education, 9*(4), 141-165.

Lee, J. H. (2001). Some comprehensive suggestions for the improvement of the College Scholastic Abilities Test (CSAT). *English Teaching, 56*(2), 333-364.

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24*(4), 517-537.

Oxford, R. (1990). *Language learning services: What every teacher should know.* Boston, MA: Heinle & Heinle.

Park, T. (2019). A comparative study of audio-mediated and video-mediated English listening tests. *Secondary English Education, 12*(3), 133-155.

Yo, I., & Rie, K. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing, 26*, 219-231.

Nayu Kim

Lecturer

Department of English Education, Jeonbuk National University

664-1 Deokjin-gu, Jeonju, Jeollabuk-do, Korea

Phone: +82-63-830-1453

Email: kimnayu98@naver.com

Heechul Lee

Professor

Department of English Education, Jeonbuk National University

664-1 Deokjin-gu, Jeonju, Jeollabuk-do, Korea

Phone: +82-63-270-2719

Email: hclee@jbnu.ac.kr