US-China Foreign Language, September 2022, Vol. 20, No. 9, 330-334
doi:10.17265/1539-8080/2022.09.003

An Assessment of English Listening Test of Senior High School Students

LI Yuhan, HUANG Fang
University of Shanghai for Science and Technology, Shanghai, China

In this study, the reliability and validity of the listening section of a sophomore monthly examination paper in a high school in Anhui Province were assessed and item analysis was conducted. The analysis revealed that the listening test has low reliability but good validity. The listening test is simple and does not have a high degree of differentiation. Most of the questions were not of high quality and the same type of questions appeared several times. In general, the listening test is not satisfactory in all aspects except validity, and the quality of the questions should be further improved to balance the reliability and validity and to improve the rationality of the listening test.

Keywords: listening assessment, reliability and validity, item analysis

Introduction

Listening tests have long occupied an important place in various types of English tests, and there have been many scholars who have studied and evaluated different types of listening tests in various ways. There have been studies on the backwash effects of listening tests on English teaching (Cao, 2009), investigations into the validity of question types and listening tests (Wu, 2001), and studies on the effects of learners' learning strategies on listening performance (Zhou, 2000). However, most of the current research has focused on studies of college English listening tests. There have been few assessments on the listening section of senior high school English tests and the current research is not in-depth. Therefore, this paper will assess the listening section of the sophomore monthly examination paper with a view to providing a reference for future content arrangement of English listening tests in senior high schools.

Literature Review

English listening is an important part of English teaching and testing and has been studied by many scholars in various ways. In terms of listening test validity, Wu (2001) studied the information processing process based on multiple-choice listening questions and found that multiple-choice questions posed a threat to the validity of the test paper. Some scholars have discussed the teaching methods of listening based on the nature and process of English listening comprehension and the factors affecting listening comprehension (Wang & Miao, 2003). Dai (2010) studied the role of task-based teaching methods in high school English listening and explored effective ways of listening teaching in high school.

LI Yuhan, Master's degree, College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China.
HUANG Fang, Ph.D., College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China.

English listening tests are conducive to the improvement of students' language communication ability and listening teaching. However, the current listening teaching in senior high school mainly based on exercises, ignoring the phased development of students' listening ability (Dong, 2017). There are still many problems in English listening teaching and listening tests in high school. Ping (2010) argued that in the stage of middle school, emphasis should be placed on developing students' listening foundation, which is essential for foreign language learning, and that English listening tests are feasible and necessary. In the future, there should be more assessment and research on high school listening tests to provide theoretical guidance on the content arrangement of listening tests and to improve the rationality of listening tests.

The Study

This study adopts a combination of quantitative and qualitative methods to assess a listening test on a monthly examination for sophomore students in a secondary school in Anhui Province. The listening test consisted of 20 multiple choice questions, each with three options and only one correct answer, with each question being scored out of two, for a total of 40 points. The first part of the test consists of five short conversations, with five questions. The second part consists of five dialogues with 15 questions. This study investigated the reliability and validity of this listening test, and an item analysis was conducted to investigate the rationality of the listening test paper.

Participants

A total of 1,080 students took this exam and a total of 117 students from two classes were selected for the study in this experiment. Class One was the experimental group, with an overall grade ranking higher than Class Two, and the actual number of participants in the exam was 61. Class Two was the control group and the actual number of participants in the examination was 56.

Data Collection

In this study, the English test answer cards of 117 participants were collected and the data from the listening test sections were entered into SPSS based on the responses on the answer cards. The data consisted of the scores for each of the 20 listening questions, the total score for each section, and the total listening score.

Data Analysis

In this study, the listening scores of 117 participants were statistically analyzed, and the reliability, validity, difficulty, and discrimination of the listening part were further analyzed.

After a descriptive statistical analysis of the 117 participants' listening scores, the following results were obtained. Table 1 show that the 117 participants scored a mean of 32.02, median of 32, mode of 36, *SD* = 5.43, minimum score of 18 and maximum score of 40, and Range = 22.

Table 1

Mean, Median, and Mode of Listening Total Score

Statistics					
Listening TS					
N	Valid	Missing	Mean	Median	Mode
117	117	0	32.02	32.00	36[a]
Skewness	Kurtosis	Minimum	Maximum		
-0.617	-0.508	18	40		

Notes. a. Multiple modes exist. The smallest value is shown.

Skewness and kurtosis are within the range of -2 to +2, which conforms to the normal distribution. The skewness value of -0.617 indicates that the majority of the participants scored above the mean, which is an initial indication that the listening test was generally easy. Kurtosis value is -0.508, which shows that the gap between the highest score and the lowest score is relatively large, which is consistent with the result of range.

Reliability

Statistical analysis was carried out after the data collection of the two sections. Through the statistical analysis of SPSS, we first obtained the overall reliability of the listening test. The reliability values for the first part and the second part of the listening test were further analyzed.

Table 2

Reliability Coefficient α of Section One and Section Two

Reliability statistics	
Cronbach's Alpha	*N* of items
0.604	6

As shown in Table 2, the reliability coefficient $\alpha = 0.604$. The overall statistical results show that, in general, the reliability of the listening test section of this paper is not satisfactory.

Table 3

Reliability Coefficient α of Section One

Cronbach's Alpha	*N* of items
0.24	5

As shown in Table 3, the reliability coefficient α for Section One was much less than 0.9. So the five questions in the first part are further counted; the first and fourth questions had a large impact on the reliability. Therefore, they were removed and further statistics were carried out, and the reliability coefficient α for Section One improved to 0.518.

Table 4

Reliability Coefficient α of Section Two

Reliability statistics	
Cronbach's Alpha	*N* of items
0.614	15

The reliability coefficient $\alpha = 0.614$ for Section Two, as shown in Table 4, which is an increase compared to Section One. After removing Questions 9 and 19, which had a negative effect on the results, and the reliability coefficient α for Section Two increased to 0.620, but the results were still below the expected value.

Validity

In this paper, the validity of the listening test was analyzed using internal correlation and the results are shown below. As shown in Table 5, the correlation between Section One and Section Two was 0.351, and the correlation between the two parts and the total listening score was 0.612 and 0.956 respectively, $p < 0.01$, which was statistically significant.

Table 5

Correlations of Section One, Section Two, and Listening Total Score

Correlations			
		TSII2	Listening TS
TSI1	Pearson Correlation	0.351**	0.612**
	Sig. (2-tailed)	0.000	0.000
	N	117	117
TSII2	Pearson Correlation	1	0.956**
	Sig. (2-tailed)		0.000
	N	117	117

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

The correlation between the two parts of the listening test was weaker, while the relationship between these two parts and the total listening score was stronger. The relationship between Section One and the total listening score is a positive weak to medium relationship, while the relationship between Section Two and the total listening score is a positive strong relationship.

Item Analysis

In the statistical analysis of discrimination index, the 117 participants were ranked according to their total listening score, with 39 participants in the top one third and 39 in the bottom one third, and 39 in the middle.

Table 6

Value of Discrimination Index

Discrimination index				
DI(N3) = 0.10	DI(N13) = 0.13	DI(N14) = 0.13	DI(N17) = 0.13	DI(N2) = 0.15
DI(N6) = 0.21	DI(N15) = 0.23	DI(N11) = 0.26	DI(N16) = 0.26	DI(N4) = 0.28
DI(N9) = 0.28	DI(N18) = 0.28	DI(N1) = 0.36	DI(N5) = 0.36	DI(N19) = 0.36
DI(N7) = 0.38	DI(N8) = 0.38	DI(N20) = 0.46	DI(N12) = 0.56	DI(N10) = 0.67
DI(average) = 0.30				

Table 6 shows the discrimination index values for these 20 listening questions. There were 12 questions with a discrimination index value less than 0.35, accounting for 60% of the total number of questions, and only Question 20, Question 12, and Question 10 had a better discrimination index value. Overall, these 20 questions were not highly discriminatory, with an average DI of 0.30, which is less than 0.35. This has a significant impact on the reliability of the listening test. Due to the limitation of the number of questions, the listening test of this paper is far below the minimum standard of reliability in terms of the discrimination of these 20 questions, which is also consistent with the results of the reliability coefficient value in the previous section.

Results

The results of the item analysis showed that the quality of most of the questions in the test itself was unsatisfactory, with 80% of the questions not having a satisfactory level of facility value. The same types of questions were repeated several times, and the content of the test was relatively homogeneous. In addition, the overall discrimination index value of the questions was less than 0.35, which did not distinguish significantly between top and bottom students and was not a good test of the differences in listening ability of students at different levels. By conducting further independent sample T-test, it was found that students with higher school

rank and lower school rank had no statistical significant difference between their listening score in this test. This also further indicated that this listening test did not well reflect the variability between the listening abilities of students with different levels of proficiency.

Conclusion

This study adopted a combination of quantitative and qualitative methods to make a detailed analysis of the listening section of the monthly high school English examination paper. The validity of the listening section was found to be satisfactory, but the reliability was unsatisfactory, the differentiation was not significant, the quality of the questions was not high, the same type of questions appeared frequently, and teachers neglected to explain the listening section in class. In the future, the balance of reliability and validity in the arrangement of the content of the listening test in high school should be paid attention to, the differentiation of the questions should be more significant, and the content of the listening test should be reasonably arranged to better reflect the listening ability of students and promote the cultivation of their listening ability.

References

Cao, Q. (2009). A study on the washback effect of CET-4 online listening test on college English teachers' teaching. *Foreign Languages and Literature, 25*(5), 164-168.

Dai, Y. F. (2010). The application of task-based teaching in senior English listening teaching under the new curriculum standard. *Modern Education Science, 27*(6), 51-52.

Dong, L. Z. (2017). A study on the washback effect of college entrance examination English listening test. *Foreign Language Testing and Teaching, 7*(1), 42-49.

Ping, K. H. (2010). The necessity and feasibility of listening test in college entrance examination. *Theory and Practice of Education, 30*(5), 17-18.

Wang, S. Y., & Miao, X. W. (2003). Theories and methods of English listening teaching. *Technology Enhanced Foreign Language Education, 25*(4), 1-5.

Wu, Y. A. (2001). Question types and the validity of listening test. *Foreign Language Teaching and Research, 45*(2), 127-135+160.

Zhou, Q. J. (2000). The influence of English listening learning strategies on listening—A survey of English listening learning strategies and analysis of the results. *Journal of PLA University of Foreign Languages, 23*(3), 62-64+95.