A Quantitative Analysis of College Listening Test: Take the Mid-term Listening Test of a College School in Xiaodian District, Taiyuan City

XUE Lei

University of Shanghai for Science and Technology, Shanghai, China

Scientific and effective exam is of great significance, for students, to optimize the learning process and grasp suitable learning methods, as well as evaluate the effectiveness of teaching and improve the quality of teaching. Therefore, based on the relevant language testing theories and with the aid of SPSS 2.0, this paper selects the listening scores from two different classes in the same grade from a college school in Xiaodian District, Taiyuan City as data, to analyze whether there is a statistical significant difference between the two classes and the underlying causes of the difference, and provides guidance for further teaching.

Keywords: quantitative analysis, listening test, college

Introduction

All examinations are designed to test the students' language performance including listening, speaking, reading, writing, and translation, of which listening is the elementary but difficult part. Zou (2005) argued, a reasonable test should pay attention to both students' receptive skills and productive skills. In order to improve teaching process, test mastery of language knowledge and the communicative ability, and provide basis for teaching reform, we have to systematically analyze these examination results with SPSS and language testing theories. Language testing, as a sub-discipline of applied linguistics, plays an important part in language teaching, and provides effective information for teaching evaluation. Based on the scientific and accurate information, teachers can adjust the teaching process. SPSS, a statistical tool, contributes to the data analysis systematically. Since 1980s, there has sprung up lots of relevant researches with the innovative ideas and methods, which affects the development of language testing. Bachman and Palmer (1981; 1982; 1988) and Clifford (1981) paid attention to the test method effect, indicating test method effect influences scores and performance of participants. Alderson (1988) was interested in the test analysis. There are lots of researches related to the characteristics of participants, and linguistic competence (Henning, 2001; Oller, 1979; Bachman, 1991; Zou, 1997).

Materials and Methods

This section introduces the materials we selected and the statistical analysis methods employed. Next, we will describe them in turn.

XUE Lei, Master's degree candidate, College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China.

Materials

In this study, the 72 listening scores from two different classes were selected as the main statistic data. Seventy-two students taught by the same teacher in the same grade participate in the examination. The paper, designed for testing students' basic listening skills, includes two parts: objective questions and subjective questions. The total score of this paper is 100 points, of which objective questions (50 points) include short conversation and extended conversation, and subjective questions (50 points) consist of completion and dictation. Because of the limitations, we get only the total scores of two classes, rather than the score of each section. Students must complete the paper within 60 minutes. The following is the scores of two different classes.

Table 1

The Scores of Two Different Classes

ID class	Class 1	Class 2
1	81.00	59.00
2	64.00	54.00
3	57.00	52.00
4	85.00	57.00
5	58.00	61.00
6	56.00	54.00
7	85.00	74.00
8	66.00	80.00
9	83.00	69.00
10	73.00	59.00
11	53.00	75.00
12	78.00	48.00
13	70.00	54.00
14	52.00	45.00
15	68.00	72.00
16	60.00	78.00
17	81.00	64.00
18	75.00	74.00
19	60.00	56.00
20	77.00	57.00
21	83.00	65.00
22	80.00	71.00
23	61.00	75.00
24	76.00	60.00
25	92.00	58.00
26	69.00	60.00
27	87.00	78.00
28	90.00	65.00
29	67.00	61.00
30	60.00	68.00
31	75.00	63.00
32	76.00	50.00
33	87.00	66.00
34	86.00	58.00
35	53.00	83.00
36	77.00	55.00

Methods

In this section, the author describes the terms related to descriptive analysis and inferential analysis respectively, which will help us analyze the data systematically.

Descriptive analysis. In statistics, descriptive analysis gives information about the general group or quality of the data, such as mean, median, and mode, as follows:

As noted by Carsten Roever and Aek Phakiti (2018), mean refers to the average value of all data; median is the value which divides the data set into two parts, so that one is larger than the median, and the other is lower than the median. Mode refers to the value that occurs most frequently in a data set. Standard deviation refers to a measure of how much the data varies around the mean.

Inferential analysis. In this thesis, we conduct an independent-samples *t*-test to analyze whether there is a statistically significant difference between two classes. Independent-samples *t*-test is used to the two different groups using the same instrument.

Test Analysis

This section we will analyze the data we collected by SPSS. Basic data analysis including mean, median, mode, range, and standard deviation, the frequency distribution of test scores, such as skewness and kurtosis, and independent-samples *t*-test will be discussed.

Basic Data Analysis

Table 2

Descriptive Statistics of the Two Classes

Statistics		Class 1	Class 2
<i>N</i>	Valid	36	36
	Missing	0	0
Mean		72.2500	63.2778
Median		75.0000	61.0000
Mode		60.00	54.00
Std. deviation		11.73122	9.65878
Skewness		-0.188	0.270
Std. error of skewness		0.393	0.393
Kurtosis		-1.165	-0.753
Std. error of kurtosis		0.768	0.768
Range		40.00	38.00
Minimum		52.00	45.00
Maximum		92.00	83.00

Note. Mean₁ = 72.25, SD₁ = 11.73, Range₁ = 40, Mode₁ = 60; Mean₂ = 63.28, SD₂ = 9.67, Range₂ = 38, Mode₂ = 54.

Based on the above data, it can be seen that the proficiency of two classes are diverse. In Table 2, the mean, median, and mode of Class 1 are higher than that of Class 2, so that students in Class 1 have a higher level of proficiency in English listening.

Concerning the variability of the scores of the two classes, we can see that the scores of Class 1 range from a minimum of 52 to a maximum of 92, while the scores in Class 2 range from a lower minimum 45 to a

lower maximum 83, indicating the scores of Class 1 are more diverse than that of Class 2. Therefore, there exists more serious two-level differentiation in Class 1.

Besides, one has a mode of 60, and the other has a mode of 54, which might not have a great effect on the data distributions. But in Class 1, the percentage of students above 60 is 83.3%, while Class 2 is much smaller, about 58.3%.

By looking at the data for Class 1 and Class 2, it is obvious that the standard deviations in Class 2 is smaller than that of Class 1, meaning the scores of Class 2 are more homogeneous and stable than that of Class 1.

The Frequency Distribution of Test Scores

From the Table 2, we can know that $skewness_1 = -0.19$, $kurtosis_1 = -1.17$; $skewness_2 = 0.27$, $kurtosis_2 = -0.75$. Language testing theory argues that skewness and kurtosis statistics should be within the acceptable range (± 2) for the assumption of normal distribution.

Based on the data, the distribution of Class1 is negative skewness with a low statistics of -0.19, suggesting there are more students with higher score. In contrast, the distribution of Class 2 is positive skewness with a low statistics of 0.27, indicating that students with a low score occupy a higher proportion. Therefore, we can say that the test is much easier for Class 1, indicating the students have a command of relevant knowledge and the proficiency with listening in Class 1 is well. In contrast, there are more students with low scores in Class 2, indicating the students in Class 2 have a low level of proficiency in English listening.

Figure 1. The histogram of the listening scores of two classes.

Figure 1 shows a distribution of listening scores of the students in both classes. In Figure 1, we can see that two curves are almost normally distributed with the help of values of skewness and kurtosis. Despite negative kurtosis, the curve of Class 1 is more flat, whereas the curve of Class 2 is a little skinny. Besides, the scores of Class 2 are more well-distributed than that of Class 1 observed from Figure 1.

Independent-Samples *t*-test Analysis

Table 3 presents the descriptive statistics of the means and standard deviations of the two classes, consistent with the Table 2. And Table 4 indicates whether the difference between two classes is statistically significant, a table with the statistics related to independent-samples *t*-test including Levene's test and *t*-test for

equality of means. In statistics, Levene's test accepts the null hypothesis that variances are equal, that is, p -value must be larger than 0.05 ($p > 0.05$) (Bachman, 2004).

Table 3

Mean and Standard Deviation of the Two Classes

Group statistics					
	Class	<i>N</i>	Mean	Std. deviation	Std. error mean
Score	201	36	72.2500	11.73122	1.95520
	202	36	63.2778	9.65878	1.60980

Table 4

Independent-Sample t-test Results

		Levene's test for equality of variances		<i>t</i> -test for equality of means						
		<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df.</i>	<i>Sig.</i> (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
									Lower	Upper
Score	Equal variances assumed	2.622	0.110	3.543	70	0.001	8.97222	2.53264	3.92103	14.02341
	Equal variances not assumed			3.543	67.512	0.001	8.97222	2.53264	3.91776	14.02669

As Table 4 shows: $F = 2.622$, $p = 0.11 > 0.05$, indicating there is no relationship between two classes' scores. Then, based on the analysis of the t -test for equality of means, it finds that there is a statistically significant difference between the scores of two classes: $t(70) = 3.543$, $p < 0.01$.

Conclusion

Through a quantitative analysis of the listening scores from two classes of a college school in Xiaodian District, Taiyuan City, we can come to some conclusions as follows:

1. The test paper is challenging. Class 1 has achieved good results, but Class 2 not. There exists a large gap between individual differences and two levels of differentiation are serious, especially Class 1. Therefore, teachers should make different plans for teaching in accordance with students' aptitude.

2. The paper's proportion is unreasonable. The paper is to test students' basic listening proficiency, in which subjective questions occupy a much higher proportion of the paper, requiring comprehensive language performance. And too many objective questions in this test maybe lead to blind guessing. Therefore, teachers should pay attention to the diversity between students and select the appropriate materials for students. Meanwhile, students should make more practice and find individual learning style to grasp the knowledge better.

3. An examination should be students-oriented, suggesting teachers should pay attention to students' performance rather than competence. According to language testing theory, competence is the basis of performance, but students should learn how to use, rather than what to know.

References

- Alderson, J. C. (1988). Testing reading comprehensive skills. In P. L. Carrell, J. Devine, and W. Grabe (Eds.), *Research in reading in a second language*. Chicago, IL: Colloquium conducted at the 22nd Annual TESOL Convention.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671-704.

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Pamler, A. S. (1981). The construct validation of self-ratings of the FSI oral interview. *Language Learning*, 31(1), 67-68.
- Bachman, L. F., & Pamler, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4), 449-465.
- Bachman, L. F., & Pamler, A. S. (1988). The construct validation of self-ratings of communicative language ability. *Language Testing*, 5(2), 260-275.
- Clifford, R. T. (1981). Convergent and discriminant validation of integrated and unitary language skills: The need for a search model. In A. S. Palmer, P. J. M. Groot, and G. A. Trosper (Eds.), *The construct validation of tests of communicative competence* (pp.62-77). Washington, D.C.: TESOL.
- Henning, G. A. (2001). *Guide to language testing: Development, evaluation and research*. Beijing: Foreign Language Teaching and Research Press.
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Roever, C., & Phakiti, A. (2018). *Quantitative methods for second language research*. New York and London: Routledge.
- Zou, S. (1997). *The Test for English Majors(TEM) Validation Study*. Shanghai: Shanghai Foreign Language Education Press
- Zou, S. (Ed.). (2005). *Language Testing*. Shanghai: Shanghai Foreign Language Education Press