

## CHAPTER FIVE

### DISCUSSION AND CONCLUSIONS

This chapter concludes the study on the accuracy of MC and SQ test formats for assessing test takers' listening comprehension. A summary of the results of data analysis presented in last chapter is followed successively by discussion of these findings, suggestions for the use of MC and SQ test formats in listening tests and the use of retrospective verbal reports, limitations of the study and suggestions for future research.

#### Summary of the Findings

The research question motivating this study was “which test format between MC and SQ best measures listening comprehension?”. Based on the results from the data analysis, generally speaking, the SQ test format was found to be better than MC test format for measuring listening comprehension. The findings showed that SQ test format gave a higher correlation between test scores and listening comprehension than the MC test format. The lower correlation between test scores and listening comprehension in MC test format was found to be caused by more frequent mismatches between student responses to test items and their listening comprehension.

The mismatches were mainly seen in the phenomena in which test takers were awarded scores when they had little or no understanding of the listening texts, and in which test takers lost test scores when they had basic or even thorough understanding of the listening texts. Two main reasons for higher frequency of the mismatches in MC test format were found: the use of test-taking strategies, and the effect of having options.

The strategies which were found to result in successfully choosing correct options with little or no understanding of listening texts in MC test format were as follows: combining what has been heard in listening texts and general knowledge to guess correct answers, matching words which have been heard in listening texts with options, eliminating impossible options and choosing from the remainders, and blind guessing without any help from listening texts or general knowledge. But these strategies could also result in failure to select the correct options. The strategy of integrating what was heard in listening texts and general knowledge to guess correct answers was found to be one of the major reasons which resulted in the failure to choose the correct option when test takers had reached basic understanding of listening texts. Another reason for the failure of choosing correct options when test takers had basic or thorough understanding was the options. The distractors in the test items were found to be misleading for test takers and disturb their choices.

It was also found in this study that MC test format can help test takers improve their listening comprehension through the information provided by its questions and options. The questions and options can give test takers the context of

listening, and make successful prediction possible which can perhaps change the listening process to match listening in real life. In addition, scoring MC questions was much easier and faster than the SQ questions.

One problem in the SQ test format was found. The results of data analysis showed that SQ format is particularly suitable for assessing understanding of listening only when the questions are clearly stated and focus on the specific content of listening texts. In the present study, it was found that when an SQ question was not clearly asking about specific information appearing in the listening text, test takers could produce incorrect answers when they had understanding of listening texts.

Taking these findings together suggests that the SQ test format has higher accuracy than the MC test format for assessing listening comprehension. In addition, these findings also contain implications for the practical use of MC and SQ test formats in assessing listening comprehension for different levels of English tests.

## Discussion

Researchers have different opinions on the use of the MC test format in language tests. Wood (1993) argues that there is no evidence to prove the existence of method effects in MC test format and that a strong correlation is found between the results of assessments with the MC test format and assessments with free response formats. However, Buck (2001) and Hughes (2003) recommend use of

SQ test format rather than MC test format to measure the understanding of clearly stated information, because designing a successful test using the MC format demands great efforts of time and expertise for preparation.

According to findings of the present study, stronger method effects were found in Listening Test 1 (MC) than Listening Test 2 (SQ), and Listening Test 2 (SQ) produced a stronger correlation between test scores and listening comprehension than Listening Test 1 (MC). The findings support the opinions of Buck (2001) and Hughes (2003).

The origins of stronger method effects of MC test format were revealed in this study. One problem which was found in the present study with MC test format was that MC questions encouraged more use of test-taking strategies than SQ questions. These test-taking strategies resulted in the inconsistency between test performance and listening comprehension in Listening Test 1 (MC). This observation lends support to a statement about test-taking strategies given by Cohen (1998). Cohen points out test-taking strategies can help test takers obtain scores for wrong reasons, but can also mislead test takers to produce wrong answers with misinterpretations of the information in options.

Another problem with MC test format in the present study was the design of options. One reason for the inconsistency between test performance and listening comprehension in the current study was the writing of options. There were some correct options which could be selected without understanding the listening texts when test takers utilized test-taking strategies successfully, and some distractors

which prompted misinterpretations when test takers had understood the listening texts. Alderson et al. (1995), Bailey (1998), Buck (2001), and Hughes (2003) all remind us that tests using MC test format demand time and effort to write. These writers suggest the following successive stages in writing MC tests: writing, analysis, adjustment, experiment, adjustment, application. However, in most situations, such stages of writing MC tests are rarely completed. At institutional level, the test designers cannot afford such a long period of preparation and heavy effort to analyze MC tests before utilizing them. Even at the national level, the participants for trialing the MC tests could be difficult to find for the secrecy requirement before the implementation of the tests. There was some evidence in this study that people who came from different backgrounds could be influenced by the same test items differently, have different testing experiences and produce different answers for the same test item. The four participants who chose the wrong option in test item 5 due to their previous learning experiences were examples of this. Therefore, the participants for trialing the MC tests should be the same people as the target group of test takers for those tests.

In addition, writing successful distractors can be difficult. Although guidance has been provided by researchers such as Alderson et al. (1995) and Buck (2001), the practice of writing options could still be a problem. According to Alderson et al, each distractor should be attractive to some of the test takers at least. Buck suggests writing distractors with words which sound similar to or have similar information to the correct option in listening tests with MC test format. But these

instructions for writing distractors are vague and could result in test designers producing MC tests in which wrong answers are potentially too attractive, leading to such misinterpretations as observed in the present study.

Researchers have also produced a formula for taking the method effects of MC test format in the consideration of scoring. Alderson et al. (1995, p. 148) and Henning (1987, p. 31-32) give the formula as follows:

$$\text{Correct score} = \text{Right answers} - \frac{\text{Wrong answers}}{\text{Number of alternatives} - 1}$$

However, they also point out that the use of this formula can only solve the problem of correcting for guessing if blind guessing is widespread and when a large number of items are omitted by test takers. The correction formula is not recommended for use in other situations. This statement actually suggests the practical limitations of the formula in most language testing situations.

Regardless of the disadvantages of MC tests, the MC test format was found in the present study to significantly promote understanding of listening texts by providing more information than SQ format through the options being previewed prior to listening to the texts. Buck (2001) believes that although opinions on the effect of question preview still stays mixed, the effect of preview as being of positive psychological value for test takers has been recommended. From the aspect of providing context for listening, the MC test format may in some ways be giving

listening experiences in tests which are more alike experiences of real life listening. Besides the advantage of providing listening context, the MC format was found easier and faster to rate.

A disadvantage of SQ test format was also found. Although SQ questions were found to have weaker method effects, were less misleading and comparatively easier to write, they were also found to be limited to measuring understanding of clearly stated and specific information. If the information is not clearly stated, or the SQ question itself is vague, test takers could be confused and produce wrong answers despite understanding of the listening texts. The answers could be hard to rate for scorers too if the question is asking for vague information, such as “what do you learn from the conversation?”.

The advantages and disadvantages of MC and SQ test formats for measuring listening comprehension are summarized in Table 11.

Table 11 Advantages and disadvantages of MC and SQ formats for listening tests

	Advantages	Disadvantages
MC Test Format	Scoring could be reliable, faster, and easier. Option preview could provide listening context to test takers.	Writing successful options could be very difficult. Test-taking strategies could have a great influence on test performance. Negative washback could greatly affect teaching and learning.
SQ Test Format	Test items could be easier to write. Test-taking strategies could have less influence on test performance. It could have higher validity as a listening test format.	Scoring could take longer and be more difficult. Questions have to be clearly stated and focus on specific information in listening texts.

#### Suggestions for Use of Test Formats and Retrospective Verbal Reports

Based on the discussion of the present study, suggestions for use focus on two aspects: future use of MC and SQ test formats in listening assessment and future use of retrospective verbal reports.

Although MC test format has been found to have stronger method effects than SQ test format in listening tests of the present study, it offers some advantages over SQ test format for assessing listening for some question types. Since the SQ test format is especially appropriate for measuring understanding of clearly stated and specific information, questions which cannot be clearly stated or do not produce specific answers can be written in MC test format to narrow down the range of possible correct answers, to promote test takers' listening comprehension, and to increase the reliability of scoring. However, due to the stronger method effects of



MC test format, the proportion of MC questions in listening tests is suggested to be reduced a minimum based on the needs of that test. It is also noted that the writing of MC questions needs caution.

The use of retrospective verbal report for investigating test takers' listening comprehension and test taking process was found successful in the present study. Bachman (1990, p. 258) emphasizes the importance of test taking process analysis. He believes that the analysis of the test taking process could be the most reliable source of information for providing insights into the factors that influence test performance. Bachman therefore recommends the usefulness of verbal reports for investigating what the language tests actually measure and facets of the method effect. Cohen (1998, p. 95) considers verbal report techniques as a major tool in collecting data on test-taking strategies, and argues for its validity as a SLA research instrument. Summing up the findings of the present study and the recommendations of Bachman and Cohen, the verbal report technique can be adopted as a useful instrument for investigating the test taking process in future research.

#### Limitations of This Study and Suggestions for Future Research

This study was conducted against the background of English testing in China. Therefore, only MC test format and SQ test format were investigated and analyzed in the present study which therefore does not provide any information on other

listening test formats. The listening text type was limited to short dialogues in order to reduce variables in the present study, hence, the influence of MC and SQ test formats on other listening text types was not investigated. In addition, due to the high demands of effort involved in the retrospective verbal report instrument, there were only 20 participants in the present study. Therefore, the results of statistical analysis in the present study might not be generalizable to other listening tests which utilize MC and SQ test formats.

The present study mainly focused on investigating the accuracy of MC test format and SQ test format on measuring listening comprehension. Some interesting directions for future research were found during the process of data collection and data analysis. One direction of future research could be investigating participants' test-taking strategies under different test response formats by utilizing verbal report techniques. Cohen (1998) suggests designing a qualitative investigation on how test takers arrived at answers to MC grammar items through retrospective verbal reports. Another suggested future study is based on the highly agreed scoring of SQ answers and understanding in the present study. Future research could investigate the rating process of scorers for marking open-ended questions.

#### Summary of the Chapter

This chapter has concluded the findings in accordance with the research

questions: SQ was found to be the better test format than MC for measuring listening comprehension. The chapter discussed the implications of these findings in response to the rationales for the study. From the discussion, suggestions for future use of MC and SQ test formats in listening tests were provided. The future use of verbal report techniques as a useful instrument for investigating test taking process was also advocated. Finally, the limitations of the present study were discussed and the suggestions for future research provided.

PAYAP UNIVERSITY