

CHAPTER ONE

INTRODUCTION

Context of the Study

Examinations are considered an important part in the educational system of China traditionally. Although in recent decades the Chinese educational system provides all children with nine years of compulsory education, six years in primary schools and three years in junior secondary schools (Liu, 1993), examinations still have a great influence in deciding whether the students can continue their education at higher levels.

After completing the studies in junior secondary schools, according to the report of education statistics produced by the Ministry of Education of the People's Republic of China (2007), only 59.2% of students continued their formal education in three-year senior secondary schools and approximately 22% of students were matriculated by colleges and universities in the year 2006. Although the chances of further education are determined by multiple reasons, the results of various entrance examinations play a crucial role in selecting students for the next level of education. The most important entrance examinations are those for senior secondary schools at the municipal level and those for colleges and universities at the national level. In addition, students in colleges and universities also have to take English tests at national level for assessment purposes which influence their graduation prospects and future employment.

The results of English examinations for selective or evaluative purposes at the municipal level or national level are all a required part of institution evaluation which is implemented by the Ministry of Education of China. Therefore, besides the functions of selecting and evaluating students, students' performance in the important municipal and national English tests also is one of the criteria for evaluating teachers and schools. This policy strengthens the influence of these tests on teaching, learning, and the form of institutional assessments.

One trend following this influence is teaching to the test format. This means that when there are some particular test formats used in these examinations, they are most likely to appear in the target classroom activities. Furthermore, communicative activities for practicing the skills are simply excluded as they do not feature in these examinations (Qi, 2003, cited in Qi, 2005). Another aspect influenced is learning. Since these tests are so important for students, students are usually inclined to choose learning materials which focus on the target test content and formats to practice after classes, and sometimes they even ask teachers to focus on what and how the target tests examine in teaching. Institutional assessments are highly influenced by municipal and national tests too. In order to raise students' scores in these tests, mock tests have become very popular in teaching and learning, and sometimes they are used as final examinations. Therefore, the test content and formats, the answer sheets, and the timetables of these tests are copied in institutional assessments.

Based on the requirement of the National English Teaching Curriculum for Full-time Primary and Secondary Schools (2001) and the National College English Curriculum Requirements (for Trial Implementation) (2003), all the municipal and national English tests emphasize the importance of listening skills. The proportion

of listening tasks in these English tests is 20% at minimum.

As a receptive skill, the listening comprehension process is invisible. It is mostly examined by reconstruction of meaning in the listening comprehension assessment tasks. However, in the process of testing, two major factors interact: one is the trait, the other is the method. Here trait refers to the listening comprehension of test takers, and method refers to the specific procedure or technique that is used to assess the trait. Test response format is one of the manifestations of test method. Since there are many different response formats to assess a given trait, they may have different method effects on the traits being measured and on the scores that test takers obtain as the result of language tests. Therefore, a good listening test is one in which the response format has little effect on the trait and students' scores are mostly a result of the trait being measured (Alderson, Clapham & Wall, 1995; Bachman, 1990; Bailey, 1998; Davies et al., 1999).

In recent years, multiple-choice format (MC) has become the most popular test format in second language listening tests in China because of the reliability of scoring. Besides that, MC format also allows scoring through computer technologies which promotes a cheaper, easier, and faster scoring process. Therefore, MC format is utilized in the listening tasks of all the important municipal and national English tests and usually used as the only format (see Appendix A which provides examples of test papers).

Influenced by these municipal and national tests, nowadays in China, assessment methods which are utilized in listening tests to measure students' listening comprehension are mostly restricted to conversations and passages with MC questions. However, Sheng (1999) and Zhou (2004) argue that MC test format has been

criticized by people for encouraging learning and teaching for successful guessing which causes strong negative washback on students for their development of real listening skills. In addition, Zhou also shows her criticism on MC test format as a measure of listening comprehension because she believes that MC could not really reveal the listening proficiency of test takers. Sheng and Zhou suggest that open-ended questions should play a role in listening assessments. Since short-answer (SQ) format is the most common open-ended test format in listening tests in China, the argument about which test format between MC and SQ can reflect test takers' listening comprehension better is naturally raised.

Statement of the Problem

As Alderson et al. (1995), Bachman (1990), Bailey (1998), Buck (2001), Cohen (1994), Hughes (2003) point out, language test takers' performance can be influenced by many variables, such as their language abilities and test response formats. In order to understand how test response formats influence test takers' performance and how well the test performance reveals test takers' actual listening comprehension, investigating the method effects of different response formats is necessary.

Researchers began to investigate test method effects in the early 1980s. Some research studies have been done on investigating if different testing methods had an effect on the test takers' scores in reading comprehension and oral ability. Research studies of this type were carried out by Alderson (1980), Bachman (1985), Klein-Braley (1983) and Shohamy (1984). Among these researchers, Shohamy

(1984) was perhaps the first researcher who focused on comparing the method effects of different testing response formats on the trait of reading comprehension and the difficulties of these formats. Inspired by these studies, researchers began to examine which of the testing methods best tapped the trait of listening comprehension and test takers' processes when taking listening tests. Buck (1991) was one of the pioneers on this type of research. He was followed by Jensen and Hansen (1995), Sherman (1997), Buck and Tatsuoka (1998), Wu (1998), Brindley and Slatyer (2002), and Cheng (2004).

However, none of these above studies answered the question which Chinese students, teachers, and test designers want to know eagerly: between MC and SQ, which test format produced test results that have a stronger relationship between the score and listening comprehension? This question is the problem addressed in the present study.

Aim of the Study

The aim of the present study is to investigate the validity of MC and SQ test formats as measures of test takers' second language listening comprehension. To this end, the method effects of MC and SQ test formats should be investigated first. Therefore, the main concern of this study is the relationship between test scores and understanding of the listening texts attained in the listening tests with MC and SQ test formats respectively. This study also attempts to explain the reasons for the results of this relationship analysis, and investigates the listening test-taking processing and the test-taking strategies. The research question that guides this study is as follows:

Which test response format between multiple-choice question and short-answer question best measures listening comprehension?

Significance of the Study

High-stakes tests usually have three major functions in countries with centralized educational systems: a selective function, an evaluative function, and an innovational function (Qi, 2005; Shohamy, Donitsa-Schmidt & Ferman, 1996). These functions mean that high-stakes tests have a great influence on teaching and learning. However, because of the conflict of different functions, negative washback could be produced as well as positive washback. In most situations in China, the municipal level and national level tests are assumed to shape teaching and learning in the way intended by policymakers and test designers, while they also have the intention to select students for higher education or evaluate students and institutions. Policymakers and test designers assume that these functions could combine with each other. Nevertheless, conflicts between the major functions of the tests make the combination very difficult.

One function of the English listening tests at national level is to produce fundamental changes in teaching and learning English listening skills according to the new national curriculums. These tests intend to lead English listening classrooms to communicative language teaching and learning. However, because of the selective function and evaluative function, teaching and learning of English listening actually turns to how to achieve higher scores in these tests. What and how the tests examine has become the main part of English listening classes. Although test content could

cover as much as that required by national curriculums, English listening tests in China mostly use MC format nowadays. Therefore, teaching and learning for successfully dealing with MC format in listening tests has become an important part of listening classes.

However, not all language teachers know thoroughly the effect of test method on the trait of listening comprehension, the importance of choosing appropriate test formats, and the influence of test formats on teaching and learning. In most situations, second language teachers act as both teachers and test designers. They are responsible for not only teaching the language but also designing and writing language tests. Therefore, the present study intends to help language teachers and language test designers to design valid listening tests.

Organization of the Thesis

The thesis is divided into five chapters. The present chapter contains an overview of the context of this study, statement of problem, aim of the study, and significance of the study. Chapter two reviews the rationales about listening comprehension and test response formats, and introduces studies on the relationship between test performance and test formats. Chapter three presents the methodology of this study which includes the explanation for the choice of retrospective verbal report as the main technique of data collection, the data collection procedure, and data analysis. Chapter four shows the results of data analysis and provides explanations for understanding the statistical data. It also provides descriptive and interpretive accounts of the data. Chapter five concludes the findings, discusses the implications

of the findings, provides suggestions for future use of MC and SQ test formats, and points out the limitations of this study and gives suggestions for the directions of future research.

PAYAP UNIVERSITY