

CHAPTER THREE

METHODOLOGY

This chapter presents the methodology of the study, including how data was collected, and how data was analyzed.

Using Introspection to Measure Listening Comprehension

The present study aims to investigate MC and SQ test formats for assessing test takers' second language listening comprehension. In order to investigate this, the relationship between test takers' scores and their understanding of listening texts has to be examined. Therefore, an accurate measure of listening comprehension with which to compare MC and SQ formats is needed.

Ericsson and Simon (1993) point out that introspective research is of a valid means enabling exploration of individuals' thought processes in task performance. Brown and Rodgers (2002) describe introspective research as that in which the researcher usually sets a task and then asks participants to report on what their brains are processing while they carry out the task. Ericsson and Simon distinguish the following three techniques in introspective research: talk-alouds, think-alouds, and retrospective studies.

According to Ericsson and Simon (1993), talk-aloud is defined as studies in which oral responses to a stimulus are concurrent with exposure to the stimulus and for which the information is already linguistically encoded and can be directly stated. Think-aloud is similar to talk-aloud but its information is not already linguistically encoded. Therefore, the information needs to be verbalized. Different from the former two types, retrospective studies produce response to the given stimuli a considerable period of time after the stimuli has been given.

Introspective studies have been conducted by many researchers for investigating human cognitive, affective, and social aspects of language learning, teaching, and testing (Brown & Rodgers, 2002). Bachman (1990) suggests that the analysis of test takers' self-reports and of actual responses should be an important part of test development because the analysis can reveal "whether the actual response is a reflection of the test takers' assertion of 'normal' language processing or of incompetence" (Bachman, 1990, p.126). Buck (1991) utilized an introspective technique to explore how listening test worked. He then recommended the further use of introspective techniques to explore listening comprehension process and reading comprehension process. Wu (1998) proved that retrospective think-aloud protocol was practical for investigating test takers' thinking process, checking their understanding, and examining the relationship between test-takers' performance in a listening test and the effect of testing response format. According to the two researchers, retrospective studies enable people to know and understand the process of listening test taking and to check test takers' understanding.

Applicability of Immediate Retrospective Technique in the Present Study

In terms of information-processing viewpoint, listening comprehension is restricted by human memory capacity. Human information processing takes some general strategies to encode, store, and retrieve information (Flowerdew & Miller, 2005). In information processing of listening input, working memory plays an important role. Working memory usually is greatly burdened when listening task demands are high because it has two main functions: storing information for later retrieval and processing the information. Therefore, during listening tests, some information processing may be slowed and some other information may be forgotten due to the high demands of tasks (Just & Carpenter, 1992). This phenomenon accounts for the fact that test takers of listening comprehension often forget what they have heard or cannot successfully construct meaning of what they have heard with their background knowledge.

Thus based on the purpose of the present study, the characteristics of listening processing, and human information capacity, talk-aloud and think-aloud may be inappropriate to be one of the data collection instruments because concurrent verbal report may hinder understanding of listening texts by increasing extra burden on working memory. On the other hand, immediate retrospective method could be an appropriate instrument to investigate participants' understanding of listening texts and their listening comprehension process because it will not cause extra burden for

working memory while the listening task is going on. For these reasons, the immediate retrospective technique was chosen for revealing test takers' understanding and their mental process when dealing with the test questions. Having established the central approach taken in data collection, the following section will outline the research procedure and give details of the data collection and analysis.

PAYAP UNIVERSITY

Data Collection Procedure

The procedure of data collection is presented in the following flow chart.

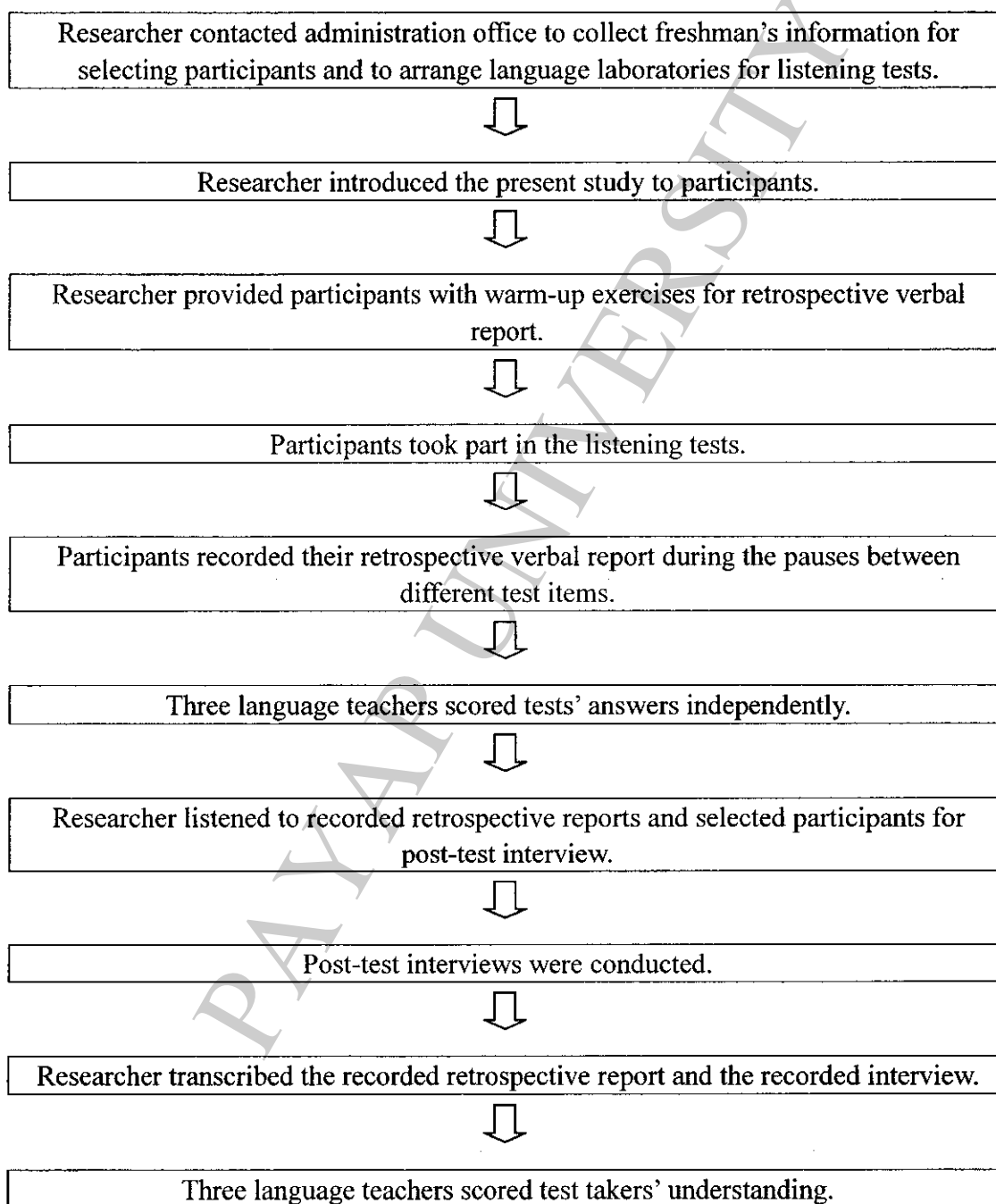


Figure 1: Overview of data collection procedure

There were 10 stages in present study for collecting data. These stages will be discussed in more detail in the following sections of this chapter.

Data Collection

Setting

For future practical use of the findings in the present study, and for cooperative and convenient purposes, the present study was conducted in the college where the researcher was working. The college is located in Southwest of China with student population of 6711 in 2006. The population of undergraduates in four-year programs was 5409, and the rest of students were in three-year programs.

English tests play an important role in Chinese students' life, and all those English tests contained listening comprehension part whose proportion was no less than 20% (see Appendix A). For this reason, English listening skills were important to every Chinese student. All of the students in this college, as an example, had taken National Matriculation English Test (NMET) before they studied in the college since NMET was a required part of National College Entrance Examinations (NCEE). All English-major students in four-year programs had to pass national Test for English Major Band Four (TEM4) at the end of their second year, and Test for English Major Band Eight (TEM8) before their graduation. Non-English major students in

four-year programs were required to achieve scores no less than 420 / 750 in national College English Test Band Four (CET4), and those in three-year programs were required to achieve no less than 60 / 100 in Practical English Test for Colleges (PRETCO).

Participants

The present study was conducted in the beginning of September in 2006 right after the college freshmen's registration. Participants were 20 freshmen who had enrolled in the English Department of the college. In order to reduce the variables, the students were selected under these criteria: same age, similar English learning experience, and similar English proficiency (see Table 1). The students were arranged into two groups and each group had 10 students. The mean of students' scores in National Matriculation English Test were the same in the two groups.

Table 1 Participants' information

Listening Test 1 (MC format)				Listening Test 2 (SQ format)			
Participant	NMET score	Age	Years of EL	Participant	NMET score	Age	Years of EL
1	109 / 150	19	6	11	109 / 150	19	6
2	109 / 150	19	6	12	109 / 150	19	6
3	108 / 150	19	6	13	108 / 150	19	6
4	108 / 150	19	6	14	108 / 150	19	6
5	107 / 150	19	6	15	107 / 150	19	6
6	107 / 150	19	6	16	107 / 150	19	6
7	105 / 150	19	6	17	106 / 150	19	6
8	105 / 150	19	6	18	105 / 150	19	6
9	105 / 150	19	6	19	104 / 150	19	6
10	104 / 150	19	6	20	104 / 150	19	6
Mean	106.7			Mean	106.7		

Note. NMET = National Matriculation English Test; EL = English Learning.

Introducing the Research to Participants

The study was introduced to the students as tests which would be used to collect information for the listening course design for freshman in the English department. The tests would be utilized to investigate English listening proficiency of the medium English level students in the classes and students' prevalent test taking processes. Students were all willing to take part in the present study and treated the study in real earnest. After the scoring process of listening tests, results of the tests were reported to the teachers of English listening of the participants to be reference for designing future course contents, and the findings of this study were reported to

the teachers later.

Listening Tests

A total of 10 contextualized dialogues were utilized to test students' listening comprehension, which were taken from the listening test of CET4 in the year of 2002 (see Appendices B and C). Since CET4 was an important national English proficiency test for only non-English major college students, the freshmen of English Department would have had no chance or desire to know it. This made CET4 an appropriate test for the present study.

Each dialogue contained the shortest possible verbal stimuli, usually a total of two turns between a male and a female native speaker of English (See Appendix C). All the dialogues, each followed with one test question, had the following characteristics: the items were content dependent and the subject matter of each dialogue contained general interest content.

Questions to the 10 dialogues were designed both in MC format with four options, and SQ format. All the dialogues were tape-recorded, the instruction, questions and options were printed on the test paper in both Chinese and English, and participants who were in the test with SQ format were required to answer questions in Chinese (see Appendix B). Therefore, there were two listening tests in the present study: Test 1 which utilized MC format and Test 2 which selected SQ format (see Table 2), with both tests sharing the same texts (see Appendices B and C and Table 2).

Table 2 Listening tests used

Listening Tests	Test 1 (Participant 1-10)	Test 2 (Participant 11-20)
Dialogue (1-10)	MC format (Dialogue 1-10)	SQ format (Dialogue 1-10)

Administration of Listening Tests

Administration of listening tests included these arrangements: warm-up exercises for accustoming the participants to retrospective verbal reports, test environment arrangement, and arrangement of retrospective report and listening test sequence.

Warm-up Exercises

Participants were given two warm-up exercises for doing retrospective verbal report and getting familiar with lab equipments before they were ready to take part in the listening tests. All the 20 participants were seated in the same language laboratory and instructed by the researcher.

Inspired by Ericsson and Simon (1993), the warm-up exercises had four steps. Step 1 was the instruction for using the apparatus in the multimedia language laboratory and practice for using it. Step 2 was the instruction for the warm-up exercises of retrospective verbal report. Step 3 was the first warm-up exercise, and

Step 4 was the second warm-up exercise.

Two exercises were selected as warm-up practice based on the research of Ericsson and Simon (1993). The first exercise was asking participants to multiply two numbers in their heads and to record what they had thought for producing the answers after they had achieved the results. The second warm-up exercise was inviting them to solve an anagram and to record their thinking process too. These two exercises were able to provide participants with consciousness of their thinking traces in their brains, therefore, helped them to produce verbal report about their thinking process and understanding.

The instruction about the warm-up exercises was guided by Ericsson and Simon (1993). It was read to participants as follows (in Chinese):

Before we turn to the real listening tests, we will start with a couple of warm-up exercises. In this exercise I am interested in what you think and what you say to yourselves as you perform two tasks that I give you. In order to do this, I will ask you to report orally after you have finished working on each of the problems and record it with your student machine while you are doing verbal report. What I want you to report is everything that you say to yourselves silently during your dealing with the problems. Do you understand what I want you to do?

OK. Firstly, I will ask you to multiply two numbers in your head and record what you think and what you say after you have done it. So multiply 45

times 37!

Good! Now I will show you a piece of paper with scrambled letters. I will ask you to find an English word that consists of all the presented letters. Any question? Record what you think after you solve the anagram.

After the warm-up exercises, participants were separated into two language laboratories according to their groups.

Test Environment Arrangement

Each group of participants were seated into a multimedia language laboratory with 120 seats which could provide all the students with same testing environment; apparatus for listening to test texts, recording verbal report, reporting to the teacher when they finished recording without interrupting other students; and comparatively far distance from each other which could ensure that they could not disturb each other while they were recording their retrospective verbal reports.

The two listening tests began at the same time when participants had finished their warm-up exercises, were seated into two language laboratories, and no more problems were reported. The tests were monitored by four teachers with two teachers in each room. One of the two teachers operated the apparatus and played the recordings while the other was distributing test paper to participants, and monitoring the whole test.

Arrangement of Retrospective Report and Listening Test Sequence

In listening tests, dialogues were played only once and after the question was asked, there were 30 seconds pause which was left for test takers to answer the questions or choose the options. Then they were required to put their answers of that question into the box on the desk which was provided to each student. Extra time was given for recording retrospective verbal report after each pause between each dialogue. The participants were asked to make detailed verbal report about what they had learnt from the dialogues and how they had dealt with the questions and options in Chinese. Each student machine in the language laboratories had a cassette recorder. After they had finished recording, they pressed calling-teacher button on the students' machine which showed a red sign on the corresponding place of teacher's computer screen. When all students had finished recording their verbal reports, the teacher gave them a sign and played the recording of next dialogue. The approximate mean time for recording retrospective reports of one test item was 3 minutes.

Scoring of Tests

Scoring was carried out on the same day as the listening tests in order to shorten the waiting period between listening tests and post-test interview, because the

longer participants waited, the less they could be expected to remember about the tests.

In order to achieve an unbiased result, the researcher did not take part in the scoring process. Three qualified and experienced language teachers were invited by the researcher to be scorers. They were colleagues of the researcher and were willing to help score the tests.

Answers of Test 1 with MC format were marked according to the standard answers which were provided by the test center of Ministry of Education in China.

In contrast to Test 1, the answers of Test 2 with SQ format were marked multiply and independently by the three scorers. In order to ensure each scorer knew thoroughly the content of dialogues and all the possible correct answers of the questions, before the scoring began, all the scorers were asked to listen to the dialogues carefully and to answer the questions by themselves. Then they were provided with consensus answers which covered all possible correct answers produced from the dialogues and discussed with each other to see if there were any differences of opinions.

The three scorers worked independently. Two scorers marked the same paper independently and recorded the results on separate score sheets, then handed them to the third scorer. When there was disagreement between the two scorers, the third one discussed with them and provided her vote.

The marking results of the first two scorers showed 92% agreement. Only eight answers among the 100 had disagreement. The third scorer rechecked and

recorded the total score of each test taker.

Post-test Interviews

Recorded retrospective reports were listened to and questions based on the report and corresponding scores of the tests were written down for post-test interviews. Only the participants whose reports were ambiguous in some aspects were selected to take part in the post-test interviews. Therefore, among these participants, only 10 of them were selected for interviews. Although there were different questions for different interviewees based on their retrospective verbal report, all of them were asked nine common questions shown in Appendix D.

Post-test interviews were held six to seven days after the listening tests. Participants were invited to the interview room separately at different times. It took two days to interview the 10 participants.

The post-test interview was designed to understand the ambiguous problems which appeared in listening to the verbal report. Therefore, the interviews were in Chinese. The tapes of the verbal report and the listening tests were played when there was necessity during the interviews to remind the interviewees of the test items and what they had reported.

The focus of this interview was mainly on two aspects: the reasons for mismatches between participants' answers and their understanding of dialogues, and their test taking strategies. The whole process of interviews were recorded and

transcribed.

Test Takers' Understanding

After post-test interviews, all the recorded data were transcribed. The recordings were transcribed to include all the hesitations, repeats, thinking pauses, corrections, and grammatical mistakes.

Then, the three teachers who had scored the listening tests were invited to mark participants' understanding of the conversations in listening tests again when the transcriptions had been done. The same multiple and independent scoring sequence was utilized as before. The recordings of retrospective verbal report were played for the teachers and the corresponding transcriptions were provided to the teachers without participants' names, their test's type, and test scores. Surprisingly, 98% agreement was produced by the first two scorers on marking the understanding, which means that there was disagreement on scoring four pieces of retrospective reports among 200 pieces. This result surprised the researcher because the agreement level was much higher than the prior expectation.

The criteria for assessing understanding of test takers were as shown in Table 3.

Table 3 Criteria for assessing understanding

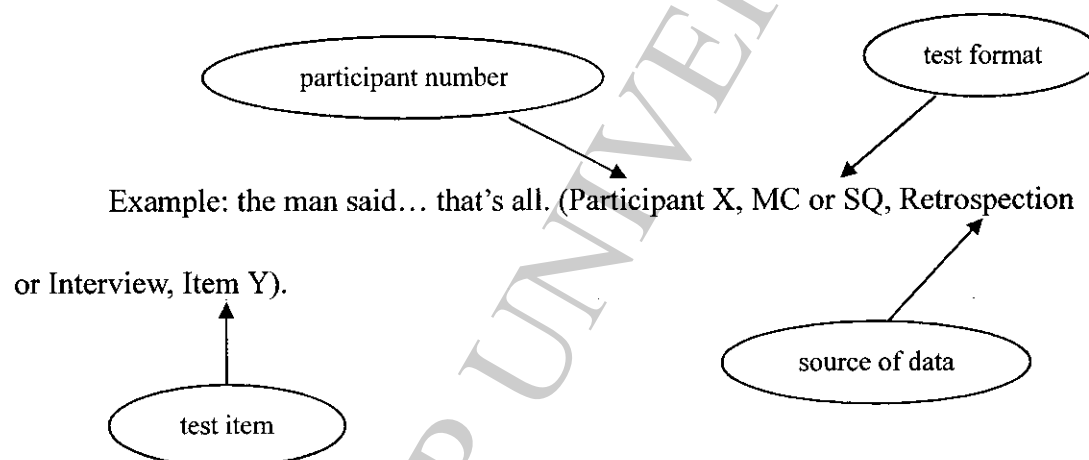
Understanding Level	Score	Description of Immediate Retrospection
Thorough Understanding	2	Test takers can understand main points of conversations in the listening test and catch important key words for answering the questions.
Basic Understanding	1	Test takers can recognize important key words in conversations for answering the questions. They can understand phrases and the highest frequency vocabulary in the listening test. They can only catch some short, clear, simple messages and announcements.
Little or No Understanding	0	Test takers can recognize familiar words and very basic phrases in conversations or they could not understand it at all. Test takers can understand basic main points of conversations but cannot catch important key words for answering the questions.

Note. The criteria were written based on the Common European Framework of Reference (Council of Europe, 2001).

Data Analysis

In order to answer the research question, the relationship between test takers' understanding of listening texts and their scores under different test formats was examined. Based on the purpose, Bivariate Correlation method was utilized to analyze the relationship between participants' scores of listening tests and their scores of understanding of the same test. Therefore, test scores obtained in MC format were compared with understanding scores achieved in MC format, and the same was done for SQ format.

Participants' viewpoints about the relationship between scores and understanding on MC and SQ formats were also analyzed through the data of retrospective verbal report and post-test interviews. For explaining how participants viewed the influence of test formats on assessing listening comprehension and their test-taking processes explicitly, the transcriptions of their verbal reports are going to be quoted in the next chapter. Each quotation of the transcriptions will be accompanied with the explanation of source as in the example:



In order to provide a complete and satisfactory explanation, each participant's scores on each test item were compared with his or her corresponding understanding on the same item. The results were categorized into three groups: test scores obtained only when participants understood the listening texts, test scores obtained when participants had little or no understanding, and test items answered incorrectly despite basic or thorough understanding.

Participants' retrospective report and the interviews provided more data for the

analysis which could support the results of statistical analysis. The results of the three groups (test scores obtained only when participants understood the listening texts, test scores obtained when participants had little or no understanding, and test items answered incorrectly despite basic or thorough understanding) which adopted MC and SQ formats were compared with each other by Independent-Samples T Test for investigating the fundamental reasons of the different method effects of MC and SQ test formats in listening tests.

Summary of the Chapter

This chapter has explained the research design, procedures, and methods used in the present study. It has provided the rationale to explain why immediate retrospective verbal report was a suitable technique for the current research. To reduce variables, the study made use of MC and SQ test formats utilizing the same listening texts in tests with the same questions, the same test environments, the same setting, and the same warm-up exercises, with tests held at the same time. In this chapter, reasons for the sequence of the present study, the arrangement of retrospective warm-up exercises have been explained. The chapter has also described how three teachers scored the answers of SQ questions and participants' understanding of listening texts multiply and independently. The following chapter will present the results of the study along with interpretation of retrospective verbal report and interviews.