CHAPTER TWO

UNDERSTANDING LISTENING COMPREHENSION AND TEST FORMATS

In order to provide more information for understanding the theoretical and empirical basis of the present study, this chapter presents theoretical notions about listening comprehension and listening test formats, and reviews studies about relationship between test performance and test formats.

Understanding Listening Comprehension

In order to investigate the relationship between listening test performance and test formats, and how well performance reflects test takers' listening comprehension, the rationales about how people understand the listening texts and obtain listening comprehension should be reviewed first. The rationales are categorized into two subsections: listening information process, and knowledge utilized in listening process.

Listening Information Process

The writers Anderson (1990), Anderson & Lynch (1988), Buck (2001), Carroll

(1986) and Rost (1994) believe that the information-processing models in listening

comprehension are bottom-up and top-down.   Moreover, the characteristics of

human memory capacity play an important role in the success of listening

comprehension.

Bottom-up and Top-down Views

Anderson and Lynch (1988), Buck (2001) and Rost (1994) point out that the

bottom-up model views listening comprehension as a linear process proceeding from

isolated units in the lower levels to higher levels step by step, and regards the output

of each stage as the input of the next higher stage.   On the other hand, top-down

model in listening comprehension stresses the influence of higher levels of

comprehension on the processing of lower levels information.

Since the two models follow opposite steps to process input information,

according to Anderson and Lynch (1988), Buck (2001) and Rost (1994), bottom-up

processing is found to utilize more phonological, lexical, syntactic and semantic

knowledge and top-down processing is found to involve the use of general

background knowledge.   The two models are parallel and interact with each other in

listening comprehension, hence various types of knowledge are involved in listening

processing in order to interpret the incoming information, and these different types of

knowledge interact and influence each other simultaneously. However, the use of

different types of knowledge can cause a burden on human memory capacity.

Human Information-Processing System View

Researchers in the psychology of language (Anderson, 1990; Carroll, 1986)

note that human information processing takes some general strategies to encode, store,

and retrieve information. In order to cope with a great amount of information

transmitted from input, three steps of human memory have to perform different

functions. The memory is categorized as sensory memory, working memory, and

permanent memory.

Flowerdew and Miller (2005) describe how these memories work to

understand listening information. Sensory memory detects signals and takes basic

treatment, encoding signals into exact form, deciding to pass them on to working

memory or discard them. Permanent memory usually assesses the information

which is committed by working memory and places the new information into its

knowledge system which has been developed, and holds the information for as long as

people wish. Permanent memory also provides the other two memory systems with

the knowledge which it saves for the analysis of information.

However, in the processing of aural information, working memory has the

greatest burden and it mainly performs two functions: storing information for later

retrieval, and processing the information. Therefore, when the task demands are high in listening comprehension tests, such as fast speaking speed or frequent unknown linguistic or non-linguistic knowledge, the computation of the brain may slow down and some partial results from working memory processing may be forgotten or lost (Buck, 2001; Flowerdew & Miller, 2005; Just & Carpenter, 1987). This could partly be responsible for the differences in difficulty between various test formats of listening tests. Zheng and Li (2002) argue that a strong correlation exists between listening comprehension and working memory based on their analysis of statistical data of two important national English tests in China.

Knowledge Utilized in Listening Process

This subsection notes the categorizations of knowledge used for obtaining listening comprehension and the importance of context in listening processing.

Types of Knowledge in Listening Comprehension Process

Anderson and Lynch (1988), Buck (2001) and Rost (1994) conclude that two types of knowledge are used in listening processing: linguistic and non-linguistic knowledge. Linguistic knowledge in listening comprehension refers to phonological, lexical, syntactic and semantic knowledge, which could provide instruction and cues for understanding the input information. Non-linguistic knowledge in listening

comprehension usually consists of general knowledge, beliefs about the world, and

the knowledge about the topic and context.

Generally speaking, linguistic knowledge is mostly utilized in bottom-up

processing.   It not only contributes to literal understanding but also facilitates the

accessing of general knowledge.   Non-linguistic knowledge is often found in

top-down processing in order to facilitate the interpretation of the incoming input.

During listening process, the two types of knowledge integrate and interact with each

other so that the deficiency of processing at any level can be compensated by the

other type of knowledge, and hence listening comprehension promoted.   Therefore,

the integration and interaction of the two types of knowledge in listening tests are

important if listening tests are to reflect this feature of listening comprehension (Buck,

2001).

Importance of Context in Listening Comprehension

In most situations, context is an indivisible part of real life communication.

It could be verbal or non-verbal in listening situations, both of which could have

considerable influence on comprehension.   For example, when a man is talking with

a woman about something important, but the woman looks at her watch several times,

this might suggest that she has something else to do at that time or she feels bored.

Therefore, the listener could predict that the man will ask if she is busy at that

moment, or the man will stop talking very soon.

However, the understanding of listening context varies among different listeners according to their language knowledge and cultural background. For this reason, providing appropriate context in listening tests is helpful for creating similar listening environments as real life by designing suitable test formats (Bailey, 1998; Buck, 2001).

Background to Study MC and SQ Test Formats

In order to understand how test response formats influence test takers' performance and how well the performance reveals test takers' actual listening comprehension, it is necessary to investigate the method effects of different response formats.

Since listening comprehension is not visible, the examiner has to rely on test takers' reconstruction of meaning in comprehension assessment tasks. Listening assessment tasks could utilize a variety of test formats (Hughes, 2003). As Buck (2001) points out, there are two macro types of test formats: selected responses and constructed responses. Selected response formats of listening tests only require test takers to produce probably a mark on a score sheet; however, constructed responses require test takers to produce some writing parts, oral answers, drawing pictures or creating diagrams. Among these response formats, short-answer questions and multiple-choice questions are the most common and widely used types in the world recently (Alderson et al., 1995; Buck, 2001; Hughes, 2003).

Multiple-Choice Questions

As a type of selected responses, multiple-choice questions (MC) with three or four options are widely used as the response format of listening comprehension. They are popular because marking can be reliable, rapid and economical; the format can make it possible to provide more items in a given period of time; tests have great reliability; and it only requires the receptive skills of test takers to answer the questions instead of production of spoken or written language (Alderson et al., 1995; Hughes, 2003).

Nevertheless, researchers (Alderson et al., 1995; Bachman, 1990; Bailey, 1998; Buck, 2001; Cohen, 1994; Hughes, 2003) also point out that this format has obvious disadvantages: writing successful items is very difficult, misinterpretations of the options might lead to the selection of wrong answers, guessing might have an effect on test scores which could threat the validity of tests, backwash might be harmful, and cheating might be facilitated.

Among these disadvantages, blind guessing is criticized the most. However, the problem cannot be solved simply by increasing options. Rogers and Harley (1999) conducted a study for investigating the method effects of three-choice and four-choice MC questions on testwiseness. They found that both of three-choice and four-choice MC questions had equal score reliability, but three-choice items lessened the influence of testwiseness.

Short-Answer Questions

The short-answer question (SQ) is a widely used type of constructed responses in listening tests.   As Hughes (2003) points out, the advantages of SQ format are as follows: items are comparatively easier to write, guessing will affect test scores less, and cheating might be more difficult.

However, this format has its own disadvantages too.   Alderson et al. (1995), Buck (2001) and Hughes (2003) think they are that responses might take longer and the possible number of items may have to be reduced, test takers have to produce language output, scoring might take longer, and scoring could be invalid or unreliable if judgment is required.

Based on the advantages and disadvantages discussed above, controversy emerges between educators on which of the two response formats could be more accurate to reveal listening comprehension in tests.   However, only a little theoretical support about this concern can be found in the literature.   Bachman (1990, p. 129) notes that constructed responses can be more difficult to test takers than selected responses.   This statement might suggest that MC format is easier than SQ format so that the test performance with MC format is better than the SQ format.   But what does difficulty mean in Bachman's statement?   Does it mean test takers can understand listening texts more with MC format than with SQ format, and this results in higher scores in listening tests with MC format?   Or is the difference due to the

effect of guessing in MC format? Does SQ format hinder listening comprehension? Based on these questions, studies were conducted in order to reveal what causes the different performance of test takers in listening tests which utilize MC and SQ test formats.

Studies on the Relationship between Test Performance and Test Format

In recent decades, the importance of listening skills has been reconsidered. Researchers began to focus on the studies of listening skills and listening assessments. The following aspects of listening assessments have been examined: task difficulty, relationship between test performance and test formats, effect of different test variables on test performance, the effect of accent on test performance, effect of authenticity on test performance, effect of stimuli in listening, and the effect of question preview on test performance.

The effect of listening test response formats on test takers' performance has been examined since the 1980s. This section presents a review of 10 studies which relate to the theme of the relationship between listening comprehension and test formats. These studies could be categorized into four groups: effect of MC, Multiple-Choice Cloze (MCC), SQ formats on listening test performance; use of introspective techniques to investigate test-taking process; method effects and task difficulty for assessing listening and reading; and effect of prior knowledge and question preview on listening comprehension.

Effect of MC, MCC and SQ Formats on Listening Test Performance

Cheng (2004) conducted an empirical study to investigate how different types of test format influenced test takers' listening performance. The test formats examined in this study were MC, MCC and SQ. She also investigated test takers' perceptions and attitudes towards these test formats.

She invited 159 technical college students to be participants and divided them into three groups for three listening tests. The same listening texts were used in the three listening tests. Each of the tests contained 10 dialogues followed by 10 questions with one of the three test formats. Post-test interviews were conducted to investigate participants' attitudes towards the target test formats. Analysis of the scores of the three listening tests revealed a statistically significant difference between the scores achieved in MC and SQ format, but no statistically significant difference was found between the scores obtained in MC and MCC test formats. The students scored higher with MCC and MC test formats. Post-test surveys showed the preference of the students to MC and MCC formats because they could use test-taking strategies to increase understanding and attain better scores, hence they were less nervous. The students also claimed that the strategies they adopted in dealing with MC and MCC test formats were mostly guessing. Cheng (2004) therefore concluded that different types of test format did significantly influence test takers' listening performance.

This study showed new orientations for future research. The results support the opinions that test takers obtain higher scores in MC format than the SQ format. However, this study provided no convincing evidence to what caused test takers' better performance in MC format, or in other words, what resulted in lower scores in SQ test format.

Use of Introspective Techniques to Investigate Test-Taking Process

Buck (1991) conducted research by using introspective techniques to investigate the listening comprehension process, how the process influenced test takers' performance, and if question preview influenced comprehension and test performance.

Participants were six Japanese students. Among them, three students took Test 1 with questions previewed and the other three took Test 2 without previewing questions. There were 13 sections in each test and test takers were asked a number of interview questions after they had completed the test items for each section. All interviewees were asked the same questions.

The results showed that the research methodology could provide valuable insights into aspects of listening comprehension processing and how the processing related to test performance, and hence the researcher recommended the future use of introspective techniques. The findings proved that SQ format had the minimal method effects on the trait but it contained three major problems: problems of

shortage of time, problems in response evaluation, and implementation problems

which could raise the unreliability of test results. Question preview was proved to

influence the choice of listening strategies, the degree of comprehension and the test

difficulty.

Inspired by Buck's work, Wu (1998) utilized retrospective verbal reports to

investigate the relationship among test-takers' performance in a listening test, their

employment of linguistic and non-linguistic knowledge, and method effects of MC

test format. The participants were 10 Chinese students who studied in Cambridge

University. MC format was the only test format utilized in the listening test. The

data collection procedure began from two pilot interviews and a pilot listening test,

followed by the real listening test. The retrospection was reported in the post-test

interviews immediately after having completed the real listening test.

The findings confirmed the necessity of using both linguistic and

non-linguistic knowledge in listening tests. MC format could provide additional

information for facilitating the use of the two types of knowledge in listening

comprehension processing with its questions and options. However, disadvantages

of MC format were found. Previewing the questions and options facilitated listening

comprehension processing for the more advanced test takers but not for the less able

test takers, misinterpretation of the options could have a partial effect on the selection

of incorrect answers, and uninformed guessing could result in choosing correct

answers for wrong reasons.

The study explicitly revealed the thinking process of test takers during the

listening test and proved the practicality of retrospective techniques for investigating

test takers' thinking process during the test and checking their understanding of

listening texts.

Rupp, Ferne and Choi (2005) conducted a study to investigate the conscious

strategies which test takers deliberately selected when they responded to MC

questions, the unconscious skills they used, and how the characteristics of passages

and MC format influenced those conscious choices and unconscious engagements.

The participants were 10 non-native speakers of English who studied in a

university in Canada.   They had a reading test with the forms of Canadian Test of

English for Scholars and Trainees (CanTEST), but all the questions were written in

MC format.   They were asked to think-aloud during responding to MC questions,

and answer some interview questions for clarification after finishing answering

questions of each section.   A questionnaire was utilized in this study too.

The findings showed that participants viewed responding to MC questions as a

problem-solving task rather than a comprehension task.   They used various strategies

to deliberately select options, such as key word matching, and combined their prior

knowledge interactively when they were choosing an appropriate option.

Summing up the findings of these studies, it is found that introspective

techniques have proved effective for investigating test-taking process, including how

test formats influenced test takers' choice of strategies to respond to test questions and

texts, how question preview facilitated comprehension of texts, and what advantages

and disadvantages MC format and SQ format had in these tests.

Method Effects and Task Difficulty for Assessing Listening and Reading

Shohamy (1984) was one of the earliest researchers who investigated the testing method effects on the trait. Her research focused on the effect of various testing methods on students' scores in language reading tests. The reading tests utilized two test formats: MC and open-ended (OE), and two languages: Hebrew (L1) and English (L2).

Participants in this study were 2000 twelfth-grade students in Israel. The controlled part was the reading test with only MC format in English; and the experimental part was the reading test containing MC questions in English, MC questions in Hebrew, open-ended (OE) English questions, and OE Hebrew questions.

The scores of the reading tests were compared statistically, and the results showed that MC format was easier than OE format; questions in Hebrew were easier than questions in English. Results proved that each of the testing response formats produced different degrees of difficulty for the test takers, and each of the variables (method, text and language) had a significant effect on students' scores of reading comprehension. These effects were strongest on low-level students.

Furthermore, the researcher suggested future research focus on investigating which reading comprehension testing method best taps the reading comprehension trait and process. The use of introspective research techniques was recommended.

Freedle and Kostin (1999) investigated construct validity of TOEFL. The

study examined the necessity for test takers to read or listen to TOEFL minitalk passages in order to answer MC questions.

The measure of item difficulty was based on the performance of approximately 2000 test takers of TOEFL.   Subjects were selected randomly from a much larger pool of test takers who had responded to each TOEFL test form. Researchers analyzed the scores of those test takers and related them to the test materials.   The results provided evidence on the construct validity of the listening minitalk items.   In addition, the researchers suggested that listening tests with MC format must demonstrate sensitivity to the information in the listening texts or else the correct options could be easily chosen by prediction or guessing.

Brindley and Slatyer (2002) investigated the effect of task characteristics and task conditions on learners' performance in competency-based listening assessment tasks, the impact on test scores of different task characteristics and task conditions, the possible reasons for differences in task and item difficulty, and the extent to which these characteristics and conditions were accounted for in the range statements.

Participants were 284 adult ESL learners in Australia and they attended three listening tests.   The data were collected only by the listening tests.   The variables investigated in this study were speech rate, text type, number of hearings, input source (live vs. audio-recorded), and item format.

The results showed that one test was easier than the other two.   However, the researchers discussed that due to the complexity of the interaction between text, speech rate, number of hearing, item and test formats, it was very difficult to isolate

the effects of specific variables from others. Therefore, the difficulty of the specific items or test formats is still unknown.

Although this study did not succeed in revealing the specific difficulty of each variable, it provided invaluable suggestions for the design of future research in this field. In order to investigate specific influence of a variable in tests, researchers might need to decrease the interference of other elements as little as possible.

Kobayashi (2002) investigated the method effects on the test performance of reading comprehension. The research mainly focused on the effects of text organization and test formats on second language learners' performance in reading comprehension tests.

Participants were 754 Japanese university students, and they were randomly divided into 12 groups. An English proficiency test was utilized as the pilot study and helped to decide the students into three different proficiency groups. After the pilot study, the reading comprehension tests which contained four text types under two different topics were implemented. The test formats were cloze, SQ question and summary writing. SQ questions and summary writing were required to be answered in Japanese and rated by qualified scorers.

The scores were analyzed, and the results showed that the differences in test performance observed across text types and test formats were statistically significant, and therefore it could be concluded that test performance was influenced by text types and test formats. Results also showed that SQ format had the strongest correlation with the proficiency test in comparison to the other two test formats. The research

also proved that text types and test formats had different influence on the students with different proficiency levels.

These studies on the method effects and task difficulty indicated that test formats did influence test takers' performance. MC test format was found easier in difficulty than other test formats, and it was also found to promote the use of test-taking strategies such as guessing and prediction. In other words, method effects were found to be strong in MC test format. SQ test format was found to have stronger relationship between test scores and reading comprehension than the test formats of cloze and summary writing.

Besides the above findings, these studies also provided invaluable references and suggestions for future research with the conclusion of their own problems and limitations.

Effect of Prior Knowledge and Question Preview on Listening Comprehension

Jensen and Hansen (1995) investigated whether prior study of a lecture topic enhanced performance on the lecture subtests of a listening test with SQ format. Their hypotheses were that listeners would need a high proficiency level to be able to use their prior knowledge, and prior knowledge could help increase listening comprehension.

A listening test which had a nonacademic section and an academic section as major components was utilized. This study also utilized self-report forms with yes

or no questions to collect information on whether the topic of the lecture was studied before.   However, the findings did not support the hypotheses. The effect size of prior knowledge was trivial. But it found the effect of listening skills was significant.

Sherman (1997) investigated the effect of question preview on listening comprehension tests.   The listening tests had four different versions, one with questions before, one with questions after, one with questions sandwiched between two hearings and one without questions.   The listening tests were written into SQ questions.

Participants were 78 undergraduates of social science at LUISS University in Rome.   The data collection instruments in this study were listening comprehension tests, questionnaires and delayed recall in written form.

Results of the tests showed no significant difference between the four listening tests in test performance although the 'sandwich' version appeared to be slightly better performed than other versions.   The results of the questionnaires did not show significant difference between the preferences of different versions either, although the 'sandwich' version came out as the most widely favored approach and the version without questions was least preferred.   The results of the delayed recall appeared to be disappointing because all the test versions failed to facilitate delayed recall of test content.   Therefore, the researcher concluded that question preview in this study had no effect on listening comprehension although it might be cognitively favorable.

These two studies showed no significant effect of prior knowledge and question preview on listening comprehension.   However, Buck (2001, p. 137)

believes the prior knowledge on the theme of listening texts and question preview

could at least provide context for listening and therefore they could have positive

psychological benefits on test takers.

Summary of the Chapter

The rationales and the empirical studies reveal that MC and SQ test formats

could have different difficulty levels which might result in different listening test

performance.   MC format was found to be easier than the SQ format and hence MC

questions help test takers attain better performance.   The different difficulty levels

could be caused by these reasons: differences in the amount of information which

could be previewed and therefore could help build listening context, and different

levels of guessing and misunderstanding caused by misinterpretation of distractors in

MC questions.

However, neither the rationales nor the research studies address the actual

listening comprehension of test takers in the listening tests with the use of MC and SQ

formats.   Which one between MC and SQ test formats could produce more accurate

results for measuring listening comprehension?

However, the empirical studies have shown orientation for the future research.

Based on the illumination of previous studies, the present study focuses on

investigating the influence of test response formats on reflecting actual listening

comprehension.   The present study avoids utilizing too many variables but just

focuses on the test format variable so that the data analysis later becomes distinct.

Instruments such as listening tests, interviews, introspective techniques are utilized in

order to achieve complete understanding of the influence of test formats on the effect

of revealing actual listening ability.   This is presented in detail in the next chapter,

which gives the research methodology of the current study.