# CHAPTER II

# LITERATURE REVIEW

This chapter presents the review of the related literature which concerns the underlying theories and framework of this study. There are three areas that are of interest in this chapter. Section One is concerned with the construct and tasks for listening comprehension tests. Factors affecting listening test performance together with test validation are also discussed in this section. Section Two presents the review of studies concerning listening tests from the last decade. The studies that reveal the effects of test task formats and English varieties are summarized in Section Two also. The studies about attitudes towards varieties of English and its study methods follow in Section Three. The review of existing literature is to portray the author's belief that there is a need to research the effects of accent varieties of English on listening comprehension ability in EFL context because there is very little research that measured the effect of English accent varieties on one's ability to understand spoken language in EFL situations. The conclusion is discussed in terms of a consideration of the design of listening test task formats and the use of English accent varieties as the listening test input.

## SECTION ONE: LISTENING COMPREHENSION TEST

### I. View on Listening Comprehension

Buck (1991) stated that "tests of second language listening ability are very common in language education, and yet a review of the literature on listening, both L1 and L2, suggests that there is no generally accepted, explanatory theory of listening comprehension on which to base these tests". This view was confirmed by Brindley (1998) in his review of assessing listening abilities that the relatively low profile of listening assessment may reflect the inherent difficulties involved in describing and assessing an invisible cognitive operation, and a number of overviews of listening comprehension have identified the lack of empirically sound models of listening comprehension which could be used to guide testing (Brown and Yule 1983, Buck, 1990, Brindley, 1998, Rost, 2002, Buck, 2001).

Among the attempts to explain listening models, Goh (2002) concluded the ideas that have influenced thinking on listening in English language teaching and he summarized these ideas into three approaches to listening, they are:

1. Listening as a skill.
2. Listening as a product.
3. Listening as a process

Goh mentioned that listening comprehension could be viewed as listening skills and proposed that the key listening comprehension skills are listening for details, listening for gist, drawing inferences, listening selectively and making predictions. Sometimes these skills are referred as 'enabling skills'.

Listening comprehension is frequently described in terms of outcomes, that is, what listeners do in order to demonstrate their understanding, and they are stated as verbal and non-verbal responses. Examples of common listening outcomes are: follow instructions, organize and classify information, take effective notes, take dictation, transfer information into graphic forms, identify information in pictures, reconstruct original text, make appropriate oral responses (Goh, 2002).

Goh's explanation of listening as a process is supported by current theory and research in cognitive science, and this leads to the main focus of this review – a cognitive model of listening comprehension.

## II. Cognitive Processes in Listening Comprehension

*Linguistic information is processed by a number of cognitive systems: attention, perception, and memory. The information is transformed in various systematic ways in the working memory and meaning is created by relating what is seen or heard to information stored in long term memory.*

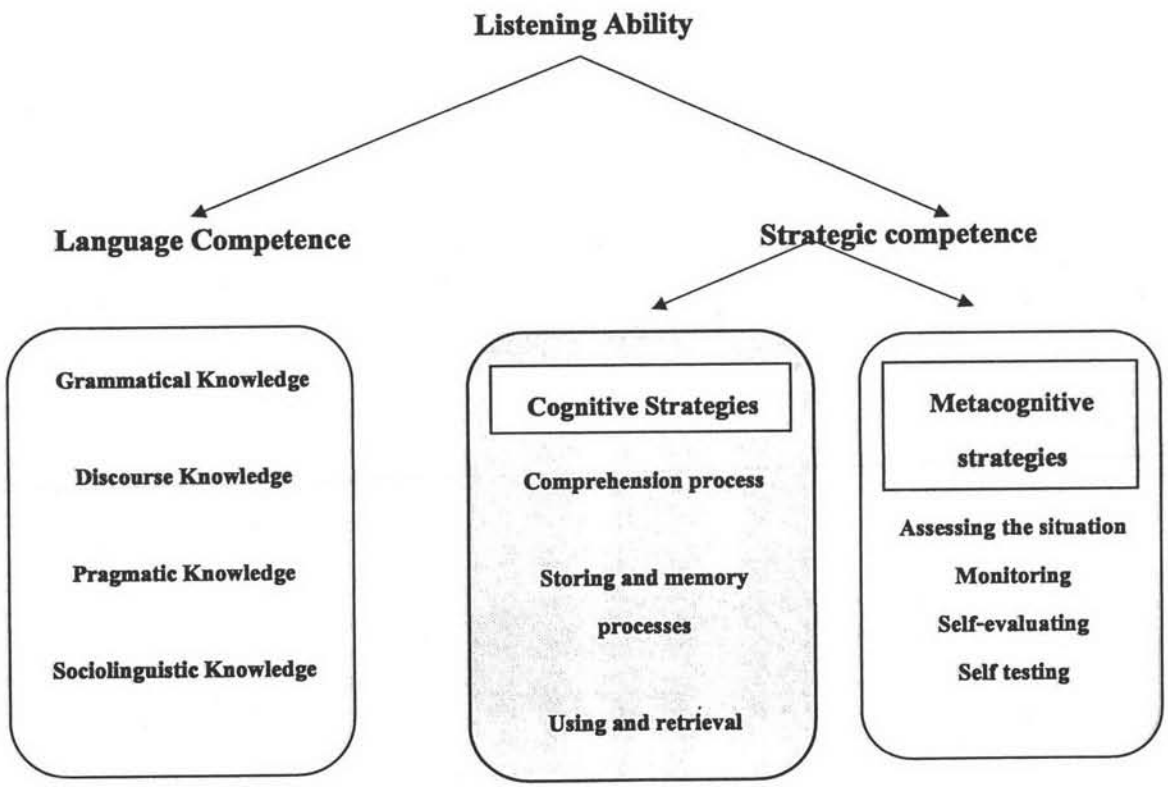
(Goh, 2002:15)

Listening is probably the least explicit of the four language skills, making it the most difficult skill to learn or assess (Vandergrift, 2004). It involves cognitive processes at different levels. Buck (1994) mentioned that most attempts to explain language processing in cognitive terms have been concerned with L1 listening, and a number of models have been proposed. Many of the earlier ones were bottom-up, serial models. Bottom-up listening refers to a process by which sounds are used to

build up increasingly larger units of information, such as words, phrases, clauses and sentences. Typically these models suggest that at the lowest level the acoustic input is decoded into phonemes, and there was an early recognition that more than one skill was involved in processing aural input.

There continues to be intense interest in the interrelationship of memory, listening and linguistic ability (Lynch, 1998). The current view is that the human being is a limited processor. According to Just and Carpenter's capacity hypothesis (1992), any listener's cognitive processes are in competition for limited processing space. A listener will have processing capacity to spare in trying to comprehend a simple text in one's own language; but in listening to an L2 text, the listener has to devote more resources to lexico-grammatical processing because of his or her limited knowledge of L2. Current models for L2 listening comprehension accept a trade-off between the storage and processing functions of working memory, with marked individual differences in listeners' skill and speed in performing operations (Lynch, 1998). A framework for describing listening ability is illustrated in the following figure.

**Figure 2.1**
**Listening Ability Framework**
**(adapted from Buck, 2001: 104)**



The fact that listeners do not rely on the text alone for comprehension is now accepted and established (Goh, 2002). There is considerable research evidence to suggest that expectations can be so strong that listeners will often hear things that were different from what was actually said. Research has shown that listeners use background knowledge (schema) to analyze, and store information. The term top-down processing is used and these models suggest that nonlinguistic skills, such as inferencing and the use of general background knowledge are an important part of listening comprehension.

Current theory and research in cognitive science clearly suggests that first language listening comprehension is multidimensional and Buck (1994) stated that there is a general assumption among L2 listening theorists that L2 processing is basically similar to L1 processing, with the obvious difference that the listener has an incomplete knowledge of the linguistic system and its use, and this is reflected in common descriptions of L2 listening (Brown and Yule, 1983; Anderson and Lynch, 1988; Rost, 1990). In the case of advanced listeners, the bottom-up processes are largely automatised. They do not need to spend time on matching sequences of sounds with written words in their mental lexicon (Goh, 2000)

Another cognitive perspective on learner listening is the use of listening comprehension strategies. Chamot (1995) stated three listening strategies: cognitive, metacognitive and social-affective. Table 2.1 explains listening strategies and their functions.

**Table 2.1**

**Listening Strategies and their Functions**

(Chamot, 1995)

| | |
|---|---|
| Cognitive | Process, interpret, store, and recall information |
| Metacognitive | Manage and facilitate mental process; cope with difficulties during listening |
| Social affective | Enlist the help of others to facilitate comprehension; manage one's emotions when listening |

This view on listening as cognitive strategies leads to tasks such as inferencing, prediction, elaboration, confidence building, attention building, comprehension monitoring, cooperation, visualizing, selective attention, and comprehension evaluation. These strategies are used to help the comprehension while listening is processed.

## III. The construct of listening comprehension tests

### a. What should be assessed?

> *Listening comprehension is a process, a very complex process, and if we want to measure it, we must first understand how that process works. An understanding of what we are trying to measure is the starting point of test constructions. The thing we are trying to measure is called a construct, and our test will be useful and valid only if it measures the right construct.*
>
> (Buck, 2000:1)

The test construct is the key issue in the testing of listening. If we are trying to assess learners' listening ability, we need to focus on those aspects of proficiency and comprehension that are unique to listening. Rost (2002) stated textual aspects and psychological aspects that are unique to listening, and they are cited as follows:

1. All physical features of spoken language that are not reflected in written language
    - pause units
    - hesitations
    - intonation
    - stress
    - variable speeds
    - variable accents
    - background sounds

2. Linguistic features that are more common in spoken language
    - colloquial vocabulary and expressions
    - shorter, paratactically organized speech units
    - false starts
    - frequent use of ellipsis
    - frequent use of unstated topics
    - more indexical expressions
    - more two-party negotiation of meaning

3. Psychological features unique to listening
    - negotiative mode: the possibility for interacting with speaker to clarify and expand meaning.
    - constructive mode: the possibility of working out a meaning that fits to context, and is relevant to the listener and to the situation, incorporating visible contextual features
    - transformative mode: the possibility of interacting with, connecting with, and influencing the speaker's ideas.

(Rost, 2002:171)

Rost (2002) insisted that these features need to be included in the listening input and the activities provided for the test takers so that test designers can be more comfortable with the construct validity of the listening test. Failure to include these features will make listening more like reading and we have to concede that we are testing an integrated set of skills that may include listening but not unique to listening (Buck, 2001).

**b. Comprehension questions**

In listening comprehension tests, the main components besides the input or the listening stimuli are the questions asked. Comprehension questions in listening may range from those which aim at assessing overall – global comprehension to those which aim to evaluate more specific – local comprehension. Shohamy and Inbar (1991) studied three questions types:

1. *global questions* that required test-takers to synthesize information or draw conclusions;
2. *local questions* that required test-takers to locate details or understand individual words;
3. *trivial questions* that required test-takers to understand precise but irrelevant details not related to the main topic.

The global questions were found harder than the local questions and the trivial questions, which relate to precise, yet irrelevant, recall of names or numerical data, were found to serve no meaningful purpose as evaluation tools. This study suggests that questions need to focus on the key information in the text, not irrelevant detail. Furthermore, the same study indicated that the combination of oral text type and local

questions was the easiest, while the combination of the literate text type and the global questions presented the most difficult test version (Shohamy and Inbar, 1991).

In 2002, Rost suggested another set of listening comprehension questions which was quite similar to Shohamy and Inbar's (1991). In his book, he proposed three kinds of comprehension questions which are 'verbatim questions', 'synthetic questions' and 'analytic questions'. Verbatim questions, which are generally the easiest, require listeners to remember specific words. Synthetic questions, which are generally more difficult, require listeners to piece together information and paraphrase ideas. Analytic questions, which are considered to be the most difficult, require listeners to analyze the meaning and draw inferences. Compared to Shohamy and Inbar's comprehension questions, these questions proposed share some similarity in that questions that require the listeners to draw conclusions and make inferences are generally perceived as harder than the questions that require memorization of specific words. This present study uses the questions proposed by Rost (2002) because one kind of Shohamy and Inbar's questions that are *trivial question* was proven to be meaningless for the test of comprehension.

**c. Listening comprehension test tasks**

Considering task characteristics brings us to the notion of interactiveness. This refers to 'the way in which the test-taker's area of language knowledge, metacognitive strategies, topical knowledge, and affective schemata are engaged by the test task (Bachman and Palmer, 1996). It means the test is testing the construct it is intended to test, and interactiveness is crucial here because it gets to the heart of construct validity of listening comprehension tests. Buck (2001) wrote that in listening we need to look at interactiveness from two perspectives: whether successful completion of the test task is dependent on comprehension of the text, and whether the knowledge skills and abilities required to comprehend the passage represent the knowledge, skills and abilities in the construct definition.

Buck (2001) suggested that listening task processing is often complex and unpredictable, some tasks are more suitable for one purpose than another. Therefore, some suggestions can be stated for test-developers to consider for their test purpose. For example, language teachers sometimes want to devise tests that focus on troublesome areas on the sounds of the target language. They can develop some

techniques for testing knowledge of the sound system suggested by Valette in 1977 (cited in Buck, 2001). These tasks include:

- minimal pairs with decontextualised words
- mimimal pairs with words in an utterance
- recognizing grammatical structures
- recognizing intonation patterns
- recognizing stress

There are tasks that are suggested also for testing understanding of literal meanings and going beyond local literal meanings. These tasks include: body movement tasks; retention tasks; picture tasks; conversation tasks; self-evident comprehension tasks; understanding gist tasks; general passage comprehension tasks; information transfer tasks (Valette, 1977 and Heaton, 1990 cited in Buck, 2001).

These tasks can be performed through a variety of test formats. A survey of testing practices done by Rost (2002) reveals several forms of listening tests as:

1. **Discrete item tests** – Multiple choice questions, open questions.
2. **Integrative tests** – Open or cloze summarizing of a listening text, complete or partial dictation.
3. **Communicative tests** - Written or oral or non-verbal tasks involving listening such as writing a letter, following direction, etc.
4. **Interview tests**
5. **Self-assessment** – learners rate themselves on given criteria, or holistic assessment of own abilities.
6. **Portfolio assessment** – learners are evaluated periodically on behavior in tasks, observations done by audio or videotapes.

Rost (2002) suggested that in order to serve as a valid measure of learners' ability, the test's input and tasks have to be consistent with what is being taught as listening. The focus of instruction tends to mirror the focus of tests. Therefore it is very important for practicing teachers, who often are test developers themselves, to use valid listening tests

## IV. Factors affecting listening test performance

A wide range of variables which may affect listening text and task difficulty has been identified by researchers. Buck (1992), Brindley (1998) and Rost (2002)

have identified several factors that influence listening test performance. Among the key variables are input factors, listener factor, and the assessment task format.

**a. The nature of the input**

Researchers have explored the characteristics of spoken texts. Among these are rate of speech, length, background, syntax, vocabulary, noise, register, and accent (Brindley, 1998). Type of text has been shown to affect understanding in listening. Brown (1995) used data from reciprocal L1 listening tasks to reveal the cognitive dimension of difficulty. According to Brown's finding, characteristics which make comprehension more difficult are 1) the number of objects spoken about, 2) their distinguishability in the text, 3) the configuration of space and time elements, 4) the explicitness of expressions used, and 5) the accessibility of the topic.

**b. Listener factors**

Comprehension of an individual very much depends on the listeners' background. These include their language ability, memory, interest, motivation and knowledge. There are views to what happens when L2 listeners tries to process L2 text input. There are two hypotheses on this issue. The earlier view is the *threshold hypothesis* which states that a listener has to achieve a certain level of proficiency in to order to process speech efficiently. The *interdependence hypothesis*, the second view, proposes that L1 and L2 listening procedural performance are similar. Zwann and Brown (1996) found evidence for both hypotheses and concluded that it depends on how one defines comprehension. From these views, it can be concluded that to be an efficient listener in L2, one needs to be efficient in L1 listening and have enough knowledge of L2 as well.

**c. The nature of assessment task**

What listeners have to do to portray their understanding of spoken text is very important in assessing listeners' listening ability. Some researchers have found that different item formats may make differing processing demands on candidates' performance in listening tests (Brindley and Slatyer, 2002). A small number of research studies have focused more specifically on particular item characteristic factors. The listening test features found in Buck and Tatsuoka's (1998) and Jensen et

al.'s (1997) works, that affects the difficulty of listening comprehension items across a range of task types include:

1. amount of lexical overlap between the text and the response format;
2. length of text preceding the information required to respond;
3. length of required response;
4. repetition of tested information;
5. whether responses and repetition of information are verbatim or paraphrases.

Furthermore, the amount of context provided, clarity of instructions, availability of question preview, type of thinking processes involved, and the use of visual context all play roles in test takers' listening comprehension ability (Rost, 2002). All of these variables need to be considered by test writers when preparing test specifications and grading listening passages as they affect the value of the test. Brindley (1998) mentioned that only a small number of studies have been conducted into role played by these variables in listening test performance. Therefore, this present study aims to fill in a deficiency in listening research by studying the effect of the nature of the input (accent varieties of English) and the nature of the assessment task (test formats). Individual attitudes (listener factors) towards accent varieties of English used as the listening stimuli is also investigated in this present research study.

## V. Listening Support on Comprehension

Much of the relevant research, which relates to the support forms incorporated with a listening test, is about the way that a listening test is administered. These listening support forms which have been studied and discussed in previous literature are previewing test questions, repetition of the input, taking notes, providing topical knowledge, and vocabulary instruction. The two popular forms which concern the present study are giving the test takers the opportunity to (1) preview the questions before they listen; and (2) to take notes while listening.

### a. Questions preview

There has been much debate about the effects of allowing test takers to preview the questions before a test begins. Some researchers (Buck, 1991; Shohamy and Inbar, 1991) suggested that question preview supplies relevant information that may orient the listeners' attention in the right direction, whereas others (Ur, 1984;

Weir, 1993) ague that question preview may alter the nature of listening processing because the prereading may distract listeners from attending to the actual input.

In an introspective study of listening comprehension processes, Buck (1991) compared three test takers who previewed questions with the other three test takers who did not. Buck (1991) found out that previewing questions

- gave useful clues about the content of the story;
- helped the test takers to listen specifically for answers;
- made the test less difficult.

According to the participants in Buck's study (1991), previewing the questions had an effect on their comprehension and influenced their choice of listening strategies. The pro-question preview was recently studied and reported in Chang and Read (2006). In their study with 160 EFL students in Taiwan, they found that question preview was effective because it helped learners to listen for necessary information and reduce the information load. However, Chang and Read (2006) mentioned the problem concerning low-language-performance test takers which was whether they could fully understand the questions and finish reading them within the time allowed. The findings from Chang and Read (2006) were consistent with the results from Sherman's study (1997). Sherman (1997) also suggested that question preview has a substantial psychological role in reducing test anxiety certainly as compared to the questions-after situation, even though the learners' actual test performance was about the same for the two conditions.

**b. Taking notes**

Hale and Courtney (1994) conducted experimental test sessions to investigate the effect of note taking on test takers' performance and their reactions to the opportunity to take notes while taking listening tests. They reported that allowing students to take notes had little effect on their performance and urging students to take notes significantly impaired their performance. These effects were observed even for students who reported being in the habit of taking many notes or reported having had classroom instruction in note- taking.

Note-taking apparently did not prove beneficial to performance, as the students' mean scores were unaffected by the opportunity to take notes (Hale and Courtney, 1994). On the other hand, one might argue that there could be no harm in

permitting students to take notes if the opportunity does not impair their performance. That is they should not be forced to take notes if they do not feel like writing notes or vice versa. This viewpoint appears to receive some support from a majority of the test takers in Hale and Courtney (1994) who indicated that taking notes helped them to answer questions better and to remember the information in the talks, and made them feel more at ease.

**VI. Test Validation**

Construct validity has become the central concern in any discussion of validity and attempts at test validation (Roever and McNamara, 2006). Messick's influential paper (1989) on the topic unifies previously separate aspects of validity – content, predictive, and criterion – under the heading of construct validity. The construct is the underlying explanation in the light of which performance on the test is interpreted. This test performance is used to draw conclusion about test takers' knowledge and their ability to perform in real world.

Messick's approach to test validation has become widely accepted in language testing and psychometrics. For example, Mislevy et al.'s work (2003) is strongly influenced by this traditional psychometric approach. The validation stages that operationalize Messick' logic could be:

- Defining the claims that score users wish to make about the test takers;
- Establishing types of evidence that are needed to support claims about test takers;
- Outlining tasks that could be used to collect necessary evidence;
- Writing the test specifications and using statistical analysis to examine claims based on evidence.

(Roever and McNamara, 2006)

Language testing research, under the influence of Bachman, has embraced the work first of Messick (Roever and McNamara, 2006). Bachman (1990) recommended the process of validation involving gathering evidence for the validity in a number of ways. His suggestions in gathering evidence are summarized as follows:

- Developing the specification of domains of content and demonstrating the tasks included in the test are representative of those specified in the domain;
- Demonstrating criterion relatedness – the steps consist of identifying an appropriate criterion behavior such as another language test or other observed language use;
- Other empirical investigation can be used such as examining patterns of correlations among different tests, observing the effects of different treatments on test scores and analyzing the process of test taking.

However, Bachman (1990) pointed out that psychometric procedures look only at test outcomes and ignore the test process itself and he proposed that evidence of test usefulness should be both qualitative and quantitative (Bachman and Palmer, 1996). There were works done in an attempt of qualitative methods such as the study conducted by Lazaraton in 2002. In her study, she went through the validation process of oral language tests without using a psychometric approach by collecting the conversation data and analyzing them (discourse analysis). From the analysis, she would had some criteria and validate the test against the criteria set. Her approach was purely qualitative and the summary was a thick description of criteria.

Furthermore, there was a work done that combined the quantitative with qualitative approach. Weir et al. (2002) discussed a detailed process of test validation that echoed what Bachman and Palmer (1996) suggested earlier to incorporate both qualitative and quantitative method in test validation. Weir et al.'s steps can be explained as follows:

*A priori validation*

**Stage 1**: Specification of the construct

The language strategies, skills and conditions might be established through:

- Target situation analysis
- Theoretical literature review
- Research literature review
- Document analysis – course book or tests

**Stage 2**: Development of pilot tests to operationalize test specification

Systematic text mapping of appropriate texts:

- To establish the consensus information recoverable according to type of language skill employed

Producing pilot test version

- Decide on most appropriate format in relation to operations
- Ensure intelligibility of rubrics
- Empirically establish timing
- Consider order of questions or process dimension
- Trial on small samples

*A Posteriori validation*

**Stage 3**: Analysis of data on the test

Trial on reasonable sample

Item analysis

Establish item

Estimates of reliability

Estimates of internal validity

Estimates of external validity

Establish what items are testing through:

- Qualitative expert judgments of items
- Qualitative introspection/retrospection by test takers
- Feedback from test takers (interview or questionnaire)

(Weir et al., 2000:5)

Weir et al. (2000) were trying to validate a proficiency test for reading in English for academic purpose. Although the present study concerns an achievement listening comprehension test, it is recommended that even achievement tests must undergo construct validation if the results are to be interpreted as indicators of ability (Bachman, 1990). Consequently, the present study makes use of this suggested process of test validation. Chapter Three concerning the study methodology will discuss the process in detail.

## SECTION TWO: STUDIES CONCERNING LISTENING TESTS

### I. Previous Studies

Reviewing the studies in language testing and assessment areas since the 1990's until present, there have not been many studies conducted that relate directly to listening tests compared to a number of studies concerning other kinds of tests in reading, writing and speaking.

Buck (1990) started the decade of the 1990s with his Ph.D. dissertation on the testing of second language listening comprehension. His thesis was the winner of the TOEFL Outstanding Dissertation Research Award in 1993. His report concluded that listening comprehension is a highly complex and individual process involving the whole of the listener's knowledge and experience, and he explored the consequences of this finding for listening test design.

In 1991, Shohamy and Inbar conducted a study which investigated the effect of both texts and question types on test taker's scores on listening comprehension tests. The listening stimuli consisted of three text types, and three question types. Results indicated that different types of texts resulted in different test scores, some kinds of listening stimuli were found easier. Also, the results showed that subjects performed better on items referring to local cues than to global cues. Implications of the results can be attributed to the selection of texts and tasks which will affect the difficulty and the construct validity of the listening tests. In the same year, 1991, Buck also reported his introspective study to examine how listening tests work. He found that listening comprehension involves much more than just the application of linguistic knowledge and he suggested that the tests should concentrate on measuring the testees' ability to understand the prepositional content of the text and should only try to test things for which the text itself provides clear criteria for judging the appropriateness of test takers' responses.

Dunkel, Henning and Chaudron (1993) proposed a framework containing range components which need to be taken into account when developing tests of listening comprehension. These include the purpose and context of assessment, the person domain (cognitive, affective) the cognitive operations required of the test takers, and the texts, tasks, items, and scoring methods used.

Later in 1994, Hale and Courtney examined the effects of taking notes in the portion of the TOEFL listening comprehension section. The results showed that allowing students to take notes had little effect on their performance and urging

students to take notes significantly impaired their performance. This might arise from the fact that TOEFL comprehension tests are designed to assess listening comprehension with minimal demand placed on memory. Buck, in the same year, reexamined the assumptions made by classical and item response measurement models and compared these with current theories of listening comprehension. The results from protocols indicate that items typically require a variety of skills for successful performance, and that these usually differ from one test taker to another. Buck stated that it is difficult to conceive of listening tests measuring one unidimensional trait on which all test takers can be placed in a linear progression from low ability to high ability, he proposed that the search for multi-dimensional measurement models, cognitive information-processing models should be done for listening comprehension tests.

In 1996, Nissan, DeVincenzi and Tang reported an analysis of factors affecting the difficulty of TOEFL listening comprehension. Five variables were found significant: 1) the presence of infrequent oral vocabulary, 2) the sentence pattern of the utterances in the stimulus, 3) the presence of negative in the stimulus, 4) the requirement of an inference for an item, and 5) the role of the speaker in the stimulus. Later in 1997, Sherman examined the effect of question preview in listening comprehension tests and reported that question preview may affect comprehension positively by focusing the attention or supplying information about the text. It can affect negatively by interfering with subjective comprehension processes, increasing the burden on the attention or imposing shallower processing. It is concluded that previewed questions seem more helpful than they really are.

In 1998, Yi'an conducted a retrospection study of Chinese EFL test-takers performing a multiple-choice task in listening comprehension tests. It is crucial to state that the MC method was found to pose threats to the construct validity of the test in two ways. First, it favored the more advanced listener, but put the less able at a disadvantage. The other way is it allowed much uniformed guessing and resulted in the subjects' giving the correct answers for the wrong reasons.

In 2002, Brindley and Slatyer reported the effects of task characteristics and task conditions on learners' performance in competency-based listening assessment tasks. The analysis of test scores suggested that speech rate and item format influence task and item difficulty, however, the complexities of the interaction between task, item, and responses suggested that simply adjusting one task-level variable will not

automatically make the task easier or more difficult. This finding confirms Bachman's posting in 2000(cited in Brindley and Slatyer, 2002) that difficulty is not a separate quality at all, but rather a function of the interaction between task characteristics and test-takers' characteristics. In the same year, Ginther studied the effects of visual condition (present or absent), type of stimuli, and language proficiency on performance on CBT listening comprehension items. The interaction between type of stimuli by visual condition, although weak, indicated that the presence of visuals results in facilitation of performance when the visuals bear information that complements the audio portion of the stimulus.

The last study to be reviewed in this part is the study conducted in 1998 by Coniam. This is the only study reported on Computer Assisted Language Learning that concerns listening comprehension directly. Coniam described the design and implementation of a computer-based listening test – the 'text dictation'. He discussed the drawbacks of classical listening test formats such as appropriate-choice selection, true-false, gap-filling, etc. that they lack what Bachman and Palmer (1996) called "interactiveness" and "authenticity". The results show that the Text Dictation program can discriminate well between students of different ability, with significant correlations obtained between students who took the computer dictation test and those who completed a traditional paper and pencil dictation test.

There are some other studies on listening tests which focus more on the method of statistic analysis, and compare listening tests with reading tests. These studies include Bae and Bachman in 1998 and Rupp, Gracia and Jamieson in 2001. Bae and Bachman's study demonstrates the usefulness of a latent variable approach to listening and reading tests. The findings show that the correlation between the listening and reading factors was high and the correlated two-factor model provided the better fit to the data, indicating that listening and reading were 'separable'. The expression 'separable' means the two factors can be separated but they share common features as well. Rupp, Gracia and Jamieson (2001) combined the techniques of multiple regression and Classified and Regression Tree (CART) analyses to study difficulty in second language reading and listening comprehension test items. The study reports that the combination of the two statistic analysis has given a richer picture of the interrelations of variables that affect item difficulty. Moreover, the findings confirm the results from previous studies that text and item interaction variables contribute strongly to item difficulty across reading and listening tests.

**II. The effect of test task formats**

The relative merits of various task formats for assessing listening were discussed in detail by a number of writers including Brindley (1998), Ur (1984), Rost (2002), Buck (2001),Goh (2002), and (Rost 1994). These response formats are varied. Ur (1984) provided a long list of the format as shown in Table 2.2.

**Table 2.2**

**Listening Comprehension Test Formats**

**(Adapted from Ur, 1984:88-9)**

| **Listening and making short responses** | **Listening and making longer responses** |
|---|---|
| • obeying instruction<br>• ticking off items<br>• true/false exercises<br>• detecting mistakes<br>• aural cloze<br>• guessing definitions<br>• noting specific information<br>• drawing pictures/maps<br>• grids filling<br>• drawing graphs and family trees | • repetition and dictation<br>• paraphrase<br>• translation<br>• answering questions<br>• predictions<br>• filling gaps<br>• summarizing |

Rost (2001) summarized that there are several ways in which listening ability is assessed. These tests can be categorized as discrete item tests, integrative tests, communicative tests, interview tests, self-assessment, and portfolio assessment.

The effect of test task on test takers' performance has been in the investigation using the psychometric approach (Yi'an,1998). In this approach, manipulations in testing methods are made and the indication of the differences which such manipulations can make is observed in the test takers' proficiency scores. Alderson (1983) and Bachman (1983) reported in Shohamy (1984), that they explored the causal factors for the differences in the test takers' scores and reported that the manipulation of different methods in cloze testing resulted in a difference in the abilities to be measured in a reading test. Shohamy (1984) expanded the scope of investigation to cover speaking through interviewing and reporting, and reading comprehension using the MC, open-ended questions, and summary testing formats. Her findings confirmed earlier results from Alderson (1983) and Bachman (1983).

Looking back through these research studies on the effect of test tasks or methods on testees' language performance, there is much room left for an investigation on the effect of test task format on listening comprehension ability of test takers, since most of the studies concerning the tasks effect have focused more on reading ability. Among the very few studies concerning the effect of test tasks on listening comprehension, Buck (1991) in his extensive research on testing listening comprehension, used the introspection method with six Japanese EFL test takers investigating the effect of open-ended short answer comprehension questions on the measurement of listening comprehension. The subjects' verbal protocols revealed the complex ways the test task influenced the measurement. The short answer method, which was found to have minimal effect in his earlier study (Buck, 1990), was reported to have some problems. His major findings include test unreliability caused by a shortage of time, response evaluation and implementation problems. As for the MC format, Yi'an (1998) also employed an immediate retrospective verbal report procedure to investigate Chinese EFL test takers performing a MC task in a listening comprehension test. From Yi'an's study, it was concluded that, while the MC format favours the advanced listener, it adds difficulty for the less able listener. Yi'an found that viewing the questions and options seemed to facilitate processing for the more advanced listener in that it helped the subject from anticipations of the incoming input and provided foci for listening, but not for the less able subjects. Furthermore, from the verbal report procedure, it was revealed that the MC format allowed much uniformed guessing. This situation led to the subjects' selection of the right answer for the wrong reason.

Moreover, certain issues of the commonly used listening test formats were discussed by Brindley (1998). He mentioned that short answer questions allow the test takers to determine fairly clearly whether the testee has understood the spoken text as long as the answers are kept very short and do not depend too heavily on test-takers' writing skills. Short answer questions also require a detailed scoring key containing a list of acceptable response, it is likely that further plausible answers will emerge during the first administrations of the test. The fact that a test item may produce such a wide range of responses makes short-answer listening questions more complicated to administer. On the other hand, true-false and multiple choice questions seem to concur short-answer questions in conditions of administration and

scoring. The obvious disadvantage of true-false is that a test-taker has a fifty percent chance of getting the right answer. Multiple-choice questions which offer ease of scoring and high internal consistency reliability are criticized on the processing demands on testees in that not only do they have to pay attention to the aural input but they also have to read and retain four propositions in working memory before matching them with the aural text.

Other formats which are commonly used in listening tests are the summary cloze and dictation. Both formats require test takers to employ others skills than listening, including memory, spelling, grammatical and lexical knowledge. It should also be pointed out that both formats are very cognitively demanding since they require test-takers to read, listen, and write simultaneously. Since this kind of tasks is very demanding, the original technique of cloze test is modified with an attempt to make them easier to such extent that Buck (2001) and Alderson (2000) agreed to call these tasks 'gap-filling'. Having test takers fill in the blanks based on what they have heard might create the most obvious problem that the test takers will not listen to the passage with true comprehension. They may not try to understand the passage at all, but may just listen for the missing words and respond based on word recognition. Brindley (1998) suggested test developers to try to include a variety of item formats in a listening test which taps a range of listening purpose.

In the Listening Comprehension Course at the University of the Thai Chamber of Commerce, these common formats – multiple choice, gap-filling, true-false and short answer, are incorporated in the course exercises that students practise in the class. These four kinds of test formats are the main tasks in the course. The present research study investigates and compares the effects of these commonly used formats in the context of EFL situation in Thailand using the psychometric approach.

**III. The effect of accent varieties of English on listening comprehension**

The global spread of English has made it an international language. The majority of uses of English occur in contexts where it serves as a lingua franca, far removed from its native speakers' norms and identities (Seidlhofer, 2001). Like other languages, varieties of English spoken in different parts of the world exhibit a certain amount of variation. Smith (1983) proposed that nonnative speakers need not sound

like the American, the British, or any other groups of native speakers in order to be effective English users. Around the world, varieties of English are spoken and variation can be seen to occur on different linguistic levels – phonetic, lexical, etc., and certainly, people are sensitive to such variation. However, in terms of listening comprehension, it is the variation of sounds that captures people's interest, especially when it is exceptional in some way. Consequenlty, varieties of English are perceived and realized as a difference in accent when people listen to speech.

Although accent is generally believed to be one of the main features that has an impact on one's ability to undertstand spoken language, very little research has measured the effects. Goh (1999) reported that 66% of learners list a speaker's accent among the factors that influence listeners' comprehension. Several studies support the widespread view that unfamiliar accents, both native and nonnative, cause difficulty in comprehension (Anderson-Hsieh & Koehler, 1988). Gynan (1985), for instance, found that listeners judged that the phonology of Spanish NNSs of English interfered with comprehension to a greater extent than grammatical errors did. In contrast, there is a good deal of evidence that familiarity with an accent aids comprehension. Eisenstein and Berkowitz (1981) reported that ESL learners more easily understood Standard English than either foreign-accented English or working-class New York English. Gass and Varonis (1984) also found that familiarity with the foreign accent was a factor in comprehension. These findings appear to be similar to Gill's study in 1994. She found that listeners with standard American accents comprehended more information when it was delivered by teachers with standard accents rather than foreign accents.

Flowerdew (1994) also suggested that learners have an advantage in listening comprehension and intelligibility when the speaker shares the same accent as the listener. This suggestion is supported by Wilcox (1988) who concluded that Singaporean learners of English found speakers of their own accent background the easiest to understand. Major et al. (2002) found that both native English speakers and ESL listeners scored significantly lower on listening comprehension tests when they listened to nonnative speakers of English. Although most researchers agree that listening to a speaker with the same variety of language as the listener enhances listening comprehension, there seems to be no consensus as to why that advantage exists or how significant its impact is (Major et al., 2005).

However, Tauroza and Luk (1997) found no support for the own accent advantage hypothesis in second language listening comprehension. They argue that familiarity with a certain foreign accent is what aids the listeners' comprehension, not the similarity between the listener's and the speaker's variety. This finding is consistent with Pihko's (1997) study, who found that Finnish and British participants found familiar Standard English varieties the easiest to understand.

A number of studies indicated that ESL learners understand standard dialects of English better than nonstandard varieties. Eisenstein and Verdi (1985) conducted a study on the intelligibility of three different social varieties - Standard English, New Yorkese, and Black English (now termed African American Vernacular English or AAVE ) - for adult learners of English. The results of their study showed that listening comprehension was indeed significantly affected by these varieties. Eisenstein in another study in 1986 found that Standard American English was easier to comprehend than New Yorkese or AAVE for working-class and middle-class immigrant ESL learners. Another study that supports the hypothesis of the effect of English varieties on listening comprehension is the study done by Sabatini in 2001. Sabatini (2001) found that nonstandard accented English varieties presented difficulties for native Italian speakers in the tasks of listening comprehension and simultaneous interpretation. In general, these findings suggest that familiarity with a given variety of English is a powerful factor in listening comprehension.

To gain a better understanding of the effect of accent varieties on comprehension, Derwing and Munro (1997) and Munro and Derwing (1999) carried out a series of experiments in which they established that comprehensibility and accentedness are related but partially independent features of nonnative speech. In these studies, they chose to obtain two types of assessments of listener comprehension in addition to foreign accent ratings. In their study, first, listeners wrote out sentences produced by nonnative speakers. They assigned scores on the basis of deviations between the transcriptions and the intended utterances. Second, they asked listeners to assign perceived comprehensibility judgements using a 9-point Likert scale. They then examined the relationships between these scores and their relationship with global foreign accent scores. Raters consistently judged accent more harshly than comprehensibility. Perhaps the most interesting finding was that although all "difficult to understand" speech samples were rated as being heavily accented, many

heavily accented samples were considered by untrained judges to be relatively easy to understand.

Another study concerning the effect of English varieties on listening comprehension was reported in 2005. Major et al. (2005) investigated whether listeners experienced more difficulty with regional, ethnic, and international varieties of English than with Standard American English variety. The participants consisted of 180 potential TOEFL takers residing in the western United States. The native backgrounds most prominently represented were Chinese, Japanese, and Spanish. All the ESL listeners had resided in the United States for 1 year or less. In addition, they included 60 native speakers of English in order to provide a baseline for comparison with nonnative speakers of English. They administered the Listening Comprehension Trial Test at six different western U.S. colleges, universities, and language institutes. The results from Major et al.'s study revealed that the interaction between speaker dialect and the listener's language status (i.e., native versus nonnative English) was not significant, but the dialect of the speaker did show a significant effect for listeners. This suggested that both native and nonnative listeners are affected by varieties of English spoken. The results demonstrated that speaker variety had a significant effect for both English as a second language listeners and native-English-speaking listeners. ESL listeners scored lower on listening comprehension tests when they heard ethnic and international varieties of English compared to Standard American English.

The latest study which concerns directly to the effect of English varieties on listeners' comprehension was conducted by Munro et al. (2006). In their study, listeners from native Cantonese, Japanese, Mandarin, and English backgrounds evaluated the same set of foreign-accented English utterances from native speakers of Cantonese, Japanese, Polish and Spanish. Regardless of native background, the listener groups showed moderate to high correlations on intelligibility scores and comprehensibility and accented ratings. These findings support the view that properties of the speech itself are a powerful factor in determining how L2 speech is perceived, even when the listeners are from diverse language background.

The results of these studies which show the effects of English varieties on listening comprehension have implications for teaching and testing English listening comprehension for L2 learners. Worldwide, most listening comprehension tests are limited to Standard American English variety or British English variety. Nevertheless, given the fact that L2 test takers encounter a wide variety of English in

real situation, it would seem judicious to include other different English varieties in teaching listening comprehension. Subsequently, testing listening comprehension with speech from the frequently heard varieties of English would seem an appropriate and valid listening ability assessment.

## SECTION THREE: STUDIES CONCERNING ATTITUDES TOWARDS VARIETIES OF ENGLISH

### I. The Method Used for Language Attitudes Studies

Higara (2005) mentioned that the so-called matched-guise techniques is the most frequently used technique for language attitude studies. This claim is proved true in many previous studies concerning attitudes (Gill, 1994; Mackey and Finn, 1997; Dalton-Puffer et al., 1997; Bayard et al., 2001, El-Dash and Busnardo, 2001). This method which was first developed by Lambert et al., 1960 (cited in Hiraga, 2005) originally aims at examining only actual language varieties and to control other variables such as the voice quality of speakers, the content of texts, or the personality of speakers. The technique requires as follows:

1. The passages are read by the same speaker who can pronounce all varieties involved in a study correctly;
2. Only one passage is used;
3. Listeners listen to the passage once and react to passage immediately.

However, it is argued that with the typical matched-guise technique, what can be measured are the attitudes towards the pronunciation of the reader of the passage or his or her accent (Hiraga, 2005). With the studies of language variations and varieties, it is recommended to use different speakers who originally represent the varieties being studied (Dalton-Puffer et al., 1997; Bayard et al., 2001). Due to the attributions of unnaturalness and artificiality, the accent authenticity of samples spoken by guises is suspicious. Because of the fact that the language varieties happen across national boundaries, it would seem almost impossible to find any person who could pronounce more than two language varieties naturally and with authenticity (Hiraga, 2005).

Usually, in the attitude studies, listeners are asked to rate their opinions on a questionnaire sheet after hearing utterances. The questionnaire sheet contains attitude traits studied. The traits investigated are power, competence, solidarity, status and voice traits. In each trait, adjective words are used for the purpose of semantic

differential scales. For instance in Bayard et al. (2001), the adjectives for the power trait were controlling, authoritative, dominant, and assertive; and the words for the solidarity were cheerful, friendly, warm, and humorous. The selection of traits for language attitude studies is varied according to the purposes of the research studies; however, the careful selection of these adjectives should be done and proved valid prior to any main investigation.

For the purposes of the semantic differential scales, before carrying out the investigation, it is necessary to decide which kind of traits – adjectives – should be chosen. Choosing the adjective words for the dimension of 'solidarity' and 'status' is not easy as Carranza and Ryan (1975) stated that:

*The ratings on the status and solidarity scales did differ somewhat; however, it is apparent that further research concerning the rating dimension is in order...What remains unclear however is, whether or not the factors (traits) obtained actually represent the activity/potency, status, and solidarity dimensions as intended by these researchers. Of particular concern is the dimension of solidarity. It may be the case that solidarity does not mean the same to the two groups involved in this particular study... ....*

(Carranza and Ryan, 1975:100)

What Carranza and Ryan (1975) mentioned is significant since it is meaningless if the factors/traits cannot represent exactly what researchers want to know. The traits that Carranza and Ryan chose for investigating the solidarity dimension were 'friendly-unfriendly', 'good-bad', 'kind-cruel', and 'trustworthy-untrutworthy'. They used 'educated-uneducated', 'intelligent-ignorant'. 'successful-unsuccessful', and 'wealthy-poor' for the status dimension. Yet it seems that the trait 'good-bad', in particular, which they included in the solidarity dimension, might fit within the status dimension because high status accents tend to be rated 'good' and 'low' status accents the opposite (Hiraga, 2005). Therefore, to obtain a degree of validity, Hiraga (2005) carried out a separate study to investigate the most suitable adjectives that reflected her subjects' language loyalties to their own languages or varieties.

Based on previous research studies concerning language attitudes (Asch,1946; Lambert et al, 1960; Levy and Dugan, 1960; Osgood, 1964; Lambert, 1967;

Strongman and Woosley, 1967; Fertig and Fishman, 1969; Tucker and Lambert, 1969; Cheyne, 1970; Giles, 1970; Giles and Powesland, 1975; Carranza and Ryan, 1975; Huygens and Vaughan, 1983; Chiba and Yamamoto, 1995) Hiraga chose 17 adjectives as candidate traits for the solidarity dimension and listed them on the questionnaire sheet. This experiment was carried out in 2002 in Oxford, UK. The 24 subjects she used were native speakers of English who were students at the University of Oxford. The reason for using native speakers of English is that the adjectives involved might have inconsistent and misleading connotations to nonnative speakers of English.

The 17 adjectives listed were: kind, friendly, reliable, conscientious, trustworthy, warm, sincere, affectionate, folksy, dependable, considerate, creative, generous, honest, sociable, religious and comforting. She listed all adjectives and asked subjects to choose from the list and also left a blank to write any adjectives that were not listed but which they thought appropriate.

The results were as follows:

(The number following each adjective is the number of subjects who chose it)

1. kind (5)
2. friendly (8)*
3. reliable (8)*
4. conscientious (4)
5. trustworthy (4)
6. warm (5)
7. sincere (7)*
8. affectionate (3)
9. folksy (1)
10. dependable (6)
11. considerate (5)
12. creative (4)
13. generous (3)
14. honest (5)
15. sociable (12)*
16. religious (1)
17. comforting (9)*

On the basis of these results, five adjectives which were rated by more than 6 subjects were chosen for the solidarity dimension. These words were sociable, friendly, comforting, sincere, and reliable.

As for the status dimension, it was less problematic with the choice of suitable adjective words. Hiraga chose the same word set from the work of Carranza and Ryan (1975). These words were educated, intelligent, wealthy, successful and elegant. Then Factor Analysis was conducted subsequent to the main study to verify that the various selected traits were clearly separated into two groups. Since there was more than one factor, they were rotated using the Varimax method. The results were as:

**Table 2.3**

**The Rotated Component Matrix of Trait Factors**

| | Component 1 | Component 2 |
|---|---|---|
| *sociable* | 0.16100 | **0.66700** |
| educated | **0.91700** | 0.07121 |
| *sincere* | 0.01569 | **0.68800** |
| intelligent | **0.84900** | 0.16700 |
| *comforting* | 0.14600 | **0.79300** |
| wealthy | **0.92100** | -0.01766 |
| *friendly* | -0.07141 | **0.82100** |
| successful | **0.81500** | 0.10700 |
| elegant | **0.89700** | 0.09239 |
| *reliable* | 0.40600 | **0.48100** |

(Hiraga, 2005:294)

From Table 2.3, ten traits were clearly divided into two factors: educated, intelligent, wealthy, successful and elegant had high loadings on one factor alone (component 1), whereas sociable, sincere, comforting, friendly and reliable had high loadings on the other (component 2). These factors were verified to be uncorrelated and they were divided into two dimensions – status and solidarity.

**II. Previous Studies on Attitudes towards English Varieties**

It is widely known that attitudes towards different varieties influence comprehension. Positive attitudes aid comprehension, while negative attitudes interfere with comprehension. Mackey and Finn (1997) emphasized that untrained native-English-speaking listeners rate nonstandard native English dialects as less natural sounding than standard speech and native nonstandard English dialects as more natural sounding than nonnative accented speech. Also, in 1994 Gill reported that native English students rated standard accents more favorably than nonnative accented speech.

In a study on language attitudes conducted a the University of Puerto Rico, Toro (1997) asked 152 students which voice they would like their English professor to have and which they would most like to imitate. Students mentioned Standard American English most favorably, followed by nonnative Greek-accented English, then Puerto Rican English, and finally Southern American English. These results suggest that an accent to which listeners have had no exposures, such as nonnative Greek-accented English, might appear more exotic and hence more attractive than more familiar accents that stereotypically have negative social associations or connotations.

ESL learners also have biases against nonnative English. Pihko (1997) found that Finnish ESL learners accepted native varieties as 'real English', whereas nonnative speech was perceived as 'strange English'. Dalton-Puffer et al. (1997) conducted a study to test attitudes towards native and non-native varieties of English in Austria. A language attitude study was undertaken with 132 university students of English. The subjects evaluated three native accents RP (Received Pronunciation), near RP and GA (General American) and two Austrian non-native accents of English. The results confirmed the low status the non-native accents have among their users and the overall preferences for the three native accents. El-Dash and Busnard (2001) who investigated Bralizian attitudes towards English also found out that Bralizian subjects in their study rated English-speaking guises more favorably than those of the native Portuguese in terms of status dimension. Not only did the students have this rigid belief in native accents but teachers of English shared the same attitudes too. In Jenkins' interview study (2005), eight NNS teachers of English were interviewed and it was reported that these teachers wanted NS English identity as expressed in a native-like accent. Native accents, according to Jenkins' study participants, were

"good", "perfect", "correct", "proficient", "competent", "original and real". Whereas nonnative accents were perceived as "not good", "wrong", "incorrect", "not real", "fake", and "deficient" (Jenkins, 2005). It seems that stereotypes regarding nonnative accented speech exist in the mind of NNSs of English, and these attitudes may have influences on listening comprehension. Generally, the respondents rate the accent best with which they have become familiar at school or during stays in English-speaking countries. However, EFL students can also have biases against some varieties of English produced by its native speakers. In 2001, Jarvella et al. investigated the language attitudes of advanced Danish students of English as a foreign language. The speech samples were all from native speakers from Ireland, Scotland, England and the USA. The speech of the Englishmen was rated as being the most pleasant of the four varieties heard, and the speech of the Americans was rated as being the least pleasant. Jarvella et al. (2001) stated that there is a more widespread positive feeling about non-American accents among young Europeans. This attitude is confirmed by another study done by Cenoz-Garcia and Lecumberri in 1999. In their study, Spanish and Basque university students rated American pronunciation less favorably not only than RP, but also than other British accents and Irish English.

On the contrary, according to Giles 1970, British people rated British regional varieties spoken in industrial conurbations such as Birmingham and Manchester much lower than American English in terms both of pleasantness and prestige. Hiraga (2005) re-examined Giles' report by using a similar technique. Sharpening the focus of Giles' study, Hiraga's experiment also found that only Standard American was significantly more favored than British regional varieties. Bayard et al. (2001) reported that the American accent seemed well on the way to equaling or even replacing RP as the most prestious or at least preferred variety in New Zealand, Australia and some non-English-speaking nations. It is, then, very interesting to investigate whether this attitude is similar for Thai listeners.

**Conclusion**

Listening is a process that we are unable to observe directly. Therefore, the construct of listening comprehension is highly elusive. It is thus very difficult to make statements about the listening ability of test takers on the basis of their test

performance. The listening comprehension ability is however assessed through a variety of item types or tasks.

Listening comprehension is a very complex process. Comprehension is subject to limitations of human memory capacity. When the task demands are high, as in a test of listening comprehension, both the storage and processing part of human working memory are needed. Some partial results from memory processing may be forgotten. This perhaps accounts for the fact that EFL listeners often seem to be able to hear everything but either forget what they hear easily and cannot process what they hear into meaning relationships (Yi'an, 1998:23). Listening comprehension is therefore seen as a part of individual cognitive ability.

When preparing a test for listening comprehension, the construct of the test should be clearly stated and the quality that is unique to listening comprehension should be included in the test. The tasks employed in the test should not make heavy demands the test takers' memory ability. The construct should focus on measuring the ability to understand the spoken text rather than the ability to remember the sound. The test, if possible, should make use of a variety of input nature and tasks in order to avoid bias against individual test taker backgrounds and differences.

Compared to a number of studies concerning others kinds of tests in reading, writing and speaking, there have not been many studies conducted directly to listening tests. Among these are the studies concerning test formats, accent and language comprehension.

Buck (2001) states native speakers are generally used to hearing a wide variety of accents. L2 listeners are usually much less familiar with the range of common accents, and they sometimes have considerable problems when they hear a new accent for the first time, and it seems to suppose that L2 listeners take much longer to adjust to a new accent than native speakers.

Accent is potentially a very important variable in listening comprehension. An unfamiliar accent can make comprehension almost impossible for the listeners. The review of the research on the relationship between accent and listening comprehension revealed that listening comprehension is a complex construct. The factors which are familiarity and degree of exposure, attitude, and stereotyping all appear to contribute to the nature of listening comprehension. In summary, listening comprehension is aided when the listener is familiar with a particular accent and has no negative attitude towards that accent.