

CHAPTER V

SUMMARY, DISCUSSIONS AND RECOMMENDATIONS

Chapter Five includes the research summary and the summary of the findings in the first part. The second part presents the discussions of the findings. Lastly, the recommendations for research consumers and future studies are provided in the final part.

I. Research Summary

The present study has the main objectives to investigate a valid construct of listening comprehension tests. The investigation focuses on two main variables that are seen to affect the test takers' ability in a test – the test method and the test input.

In listening comprehension tests, two main components interact – the listening stimuli and how test takers perform answers. Most of the listening tests found, whether they are a high stake test or a tailored made test, the use of the multiple choice method with native speakers voice stimuli are common. In an era of international English, the question of whether the test will reflect authentic use of English is of interest. Moreover, the use of the multiple choice format per se in most test situations is questionable, especially in listening comprehension tests when simultaneous interpretations are needed for listeners. The study concerning these two variables could help shed the light on this dilemma.

It is widely known that attitudes towards English varieties play an important role in language learning and comprehension. People usually react differently when they hear varieties. They might like or dislike the accents heard. It is very interesting to investigate these attitudes and preferences of test takers as this can be a link to help explain their performance on the listening comprehension test that makes use of English accent varieties as voice input. All of these interests lead to four research questions in this present study. They are:

1. Do different test formats have a significant effect on the test takers' listening comprehension ability? And if they do, how much is the effect size?
2. Do English accent varieties used in listening input have a significant effect on test takers' listening comprehension ability? And if they do, how much

is the effect size?

3. Is there a significant interaction effect between different listening test formats and accent varieties of English? And if there is, how much is the effect size?
4. What are the test takers' attitudes towards accent varieties of English and their preferences towards using accent varieties of English as listening comprehension verbal input?

The two main variables were studied concerning their effect on the test takers' scores. Because these two factors were used together in the test (the test method with the test input) their interaction effect was also investigated. Further, the test takers were interviewed to explore their attitudes and preferences on English accent varieties to find support to the experimental research findings

Selecting the test format and the English accent varieties to be studied

The two main variables which were the test method and the test input were selected to be studied concerning their effects on the listening comprehension scores of test takers. For the test method, the format of multiple choice and short answer were compared; and for the test input, native varieties and nonnative varieties of English were examined in contrast. British, American, and Australian English were included in native speakers' varieties; on the other hand, nonnative accent varieties were those from Malaysia, Japan, China, and Singapore. The rationales for selecting these independent variables were:

1. A pilot study was conducted to select the appropriate test format to be used in the main study. Since the test constructed was an achievement test for the course taught at UTCC, the test formats commonly used in the classroom which were multiple choice, gap filling, true-false, and short answer were included to be investigated in the pilot study first. The findings show that the multiple choice format was significantly different from the other three formats, and so was the short answer. While gap-filling test scores are not different from true-false test scores, and gap-filling in the pilot study was proved to be easy and did not truly evaluate test takers' comprehension. Because, in the pilot study, gap filling and true-false test format produced similar results, consequently, only multiple choice and short answer were selected to

be investigated further in the main study. However, the researchers assumed that all dialogues, questions, and exercises used in the course were at about the same level of difficulty. Therefore, the difficulty level of the four test formats was not strictly controlled in the present study.

2. There are a lot of English varieties used around the world. However the experimental study had to limit the number of accents to be tested, it was therefore reasonable to focus on those that the test takers will have more chance to experience. According to the statistics reported by the two government offices –BOI and TAT in 2005, the people from the countries that have visited and invested the most in Thailand are ranked as follows: 1) USA, 2) UK and 3) Australia for the native English speaking countries; and for the nonnative English speaking countries they are from 1) Japan, 2) Malaysia, 3) China and 4) Singapore. The English varieties from these countries are therefore included in the present study.

The listening comprehension test

An achievement test of listening comprehension for a course called 'Listening Comprehension' was developed to investigate the two main factors mentioned. The test went through a validation process by asking five course lecturers to look at content, dialogue and questions in the test. The students, who studied in the course, were also asked questions concerning the test formats used and the test content. Finally, statistic validation such as reliability index, difficulty index and item analysis were employed to make the most possible valid test for the main study. The test was improved and then transformed into test versions A and B - the native speaker version and the nonnative speaker version respectively. The selection of the speakers' voice was controlled only minimally using the criteria as education level, age, gender and language fluency.

Participants

The participants were 192 second-year students from the School of Humanities, University of Thai Chamber of Commerce, majoring in English for Business Communication. They were studying Listening Comprehension in the

second semester of 2005 academic year. The 192 students were the sample from the total population of 380 students. In the interview phase, only 30 students selected from 192 participants were invited for the study.

Instruments and data analysis

The research study instruments consisted of the listening comprehension test version A (native varieties) and version B (nonnative varieties). The semi-structure interview questions were used to study the attitude and preferences of the test takers towards varieties of English. In terms of statistical procedure, a two-way mixed Factorial ANOVA was used to investigate the main and interaction effects of two variables on the test scores. The measure of effect size was carried out by using the suitable formula of the Cohen's *d* for *F* test ratio, and the Eta squared to report the size of experiment effect. To examine the test takers' attitudes towards varieties of English, a modified matched-guise method, in which the participants were asked to listen to the seven varieties and react to them by giving their opinions on each voice, was employed. The mean scores of attitude scales were calculated to show the average scores of the seven varieties. The qualitative data from interview questions enquiring about the participants' preference for using English varieties was reduced, counted, categorized, and displayed in the form of tables and graphs. Conclusion drawing and verification were done by consulting two more judges for reliability check.

Findings

1. The test takers' scores from the multiple choice format were significantly different from the scores provided by the short answer format at .05 level. The results from the ANOVA illustrates that there was a significant main effect for the test formats, $F(1, 190) = 25.62, p < .001$, with a medium effect size of $\eta^2 = .119$. This suggests that the test-takers' listening ability is affected by the test formats used. Moreover, the mean score for the short answer format is higher than the mean score from the multiple choice from both test versions; therefore, it proves that the multiple choice format is more difficult than the short answer format for the test takers.

2. Using different varieties of English had a significant effect on the test takers' scores. The test version A, which used native-speakers' voice, was easier than

the test version B that used nonnative-speaker voice for the test takers in this present study. The statistic results, which are $p < 0.05$, with a small effect size of $\eta^2 = .021$, show that although using English varieties was proven to have a significant effect on the scores, the size of experiment effect was very small.

3. There was a significant interaction between the test formats and varieties of English, $F(1, 190) = 6.30, p < 0.05$, with a small effect size of $\eta^2 = .032$. The interaction is significant at the level of .05. Moreover, from the graph plotted, the pattern of interaction indicates that there was virtually not much difference between the scores in the test version A and B for the multiple choice format (test A mean = 11.343, SD = 3.996 ; test B mean = 11.020, SD = 3.283), and from the *t*-test, the difference between the two versions was not significant. The multiple choice format variable did not produce much difference in scores in both test versions. It is clear from the interaction pattern that the combination of the short answer format with native varieties of English is the easiest test version, whereas the multiple choice formats that comes with nonnative speakers' voice is the hardest test version for the test takers.

4. For the test takers' attitudes towards English varieties, it can be summarized as follows:

- (1) The test takers have a more positive attitude towards the native varieties than the nonnative varieties of English;
- (2) American English is the most acceptable variety in both solidarity and status aspects. Second to the American accent is British English with a very close mean score to American English of 2.94 and 3.10 respectively. This is well supported by the interview questions about the respondents' preferences for the English accent, 9 respondents mentioned the American variety and 8 interviewees mentioned British English.

As for the preferences for the varieties used in the listening comprehension test, it can be summarized as:

- (1) 80% of the respondents agreed that nonnative varieties are harder to comprehend when compared to its counterpart – native varieties;
- (2) The majority of respondents (86.7%) agreed that the test which includes nonnative varieties voice would be considered more difficult for them;
- (3) Only 33% of respondents indicated that they want English varieties from nonnative speakers to be included in the test while the rest prefer only

native speakers' voice;

- (4) More than half of the respondents (53.3%) are not uncomfortable to listen to accent varieties of English used in the test. Particularly, listening to English varieties is more preferred in a non-test situation.

II. Discussions

The purposes of this study were to investigate the effects of two main variables which were 1) test formats and 2) varieties of English on the listening comprehension test scores. The interaction between the two main variables was also examined. The test takers' attitudes and their preferences towards accent varieties of English were investigated through semi-structured interviews.

There were altogether four research questions. Three hypotheses were set to correspond to the first three quantitative research questions. The fourth qualitative question was answered by semi-structured interviews. The discussions based on the findings related to the hypotheses and questions are presented in the following section.

Question 1

“Do different test formats have a significant effect on the test takers' listening comprehension ability? And if they do, how much is the effect size?”

Hypothesis 1

“There is a significant difference in the scores obtained from the multiple choice format and the short answer format.”

To test hypothesis 1, if there is a significant difference in the scores obtained from the multiple choice format and the short answer format, a mixed between-within ANOVA was conducted. It was found that the scores from the multiple choice format are significantly different from the short answer format scores at the 0.05 level. Concerning the effect size value, the Cohen's *d* effect size value of 0.73 which is close to a large effect size level indicates that employing different formats in the listening comprehension test had quite a strong effect on the participants' scores. The results from the main study also confirm the pilot study results that employing different test formats affects the test takers' scores.

Therefore, it can be concluded from these findings that using different test formats in the test of listening comprehension definitely has an effect on the test takers' ability. The findings lend support to the study done by Alderson (1983) and Bachman (1983) that manipulation of different methods resulted in a difference in the abilities to be measured in a reading test. Shohamy (1983), who expanded the scope of investigation to cover the speaking skill, also found out that using different test format (MC, open-ended questions, and summary testing formats) affected test takers' speaking scores. The results from this study not only confirm the earlier works of Alderson (1983), Bachman (1983) and Shohamy (1984), but also expand the scope to cover the listening skill. It can be summarized that using different test methods in a test of listening comprehension can produce different results of test takers' scores in a similar manner to a test of reading and speaking.

Furthermore, it can be seen from the test results that although the test takers prefer to have the test in the form of multiple choice because the format allows them for guessing, the study results, surprisingly, suggest that the listening comprehension test with multiple choice format was more difficult than the test with the short answer format for the test takers.

However, the findings found in the present do not concur with the findings from Berne's study (1995). Berne's findings showed that test takers performed better on multiple choice questions than on either an open-ended or cloze task. Brindley and Slatyer (2002) explained that items requiring only cognition are easier than those requiring retrieval and production. They mentioned the lack of study of effects of short answer format questions. The difference between Berne's study and this research study is that in Berne's research study, most questions required longer answer than the present study did. On the other hand, in the present research study, the test takers were required to give a very short answer of not more than two or three words, and a lot of items needed only one word to be marked as a correct answer. This might be the cause for the test takers to find short answer items quite easier when compared with multiple choice items. The present study followed what was mentioned by Brindley (1998) in his extensive review of assessing listening abilities that short answer questions can work quite well in listening tests, provided that the answers are kept very short and thus do not depend too heavily on candidates' writing skills. This might be the cause that the short answer format, in this present study situation, was found to be easier than the multiple choice format.

Why was the test takers' average scores obtained from the short answer format different and higher than those from the multiple choice format?

The average scores from the short answer format were found to be different and higher than the average scores from the multiple choice format. The effect size of the test format was quite large ($d = 0.73$). This could be due to the following reasons:

1. Successful listening comprehension involves a number of different skills and an array of knowledge – some are linguistic knowledge and some are nonlinguistic. Nonlinguistic knowledge, such as inferencing and the use of general background knowledge, is also important in listening comprehension. The protocols in Yi'an's study (1998) and Buck's study (1994) clearly indicate that listening comprehension is a process of making sense of the linguistic input in the light of nonlinguistic knowledge. In work on language comprehension, the most frequently discussed nonlinguistic variable is background knowledge (Buck, 1994). In a multiple choice listening comprehension test situation, the questions and optional answers serve as the imposed and therefore shared purposes for listening and exert an impact on the test takers' listening processes (Yi'an, 1998). In the present study situation, the test takers were listening to the stimuli and questions while reading the multiple choice options; this forced the subjects to activate relevant general knowledge in the light of options concerned in order to make sense of what was heard and read at the same time. In this case, linguistic and nonlinguistic knowledge was used by the test takers to understand the stimuli, questions and options. Nonlinguistic knowledge was accessed to compensate for the test takers' linguistic deficiency in order to cope with the task (Brown and Yule, 1983), although such compensation did not guarantee arriving at the correct answer. In performing the MC task, the test takers were using linguistic and non-linguistic knowledge in order to understand the spoken text and the written text that was used in options. The length of the MC task was definitely longer than in the short answer task in the present study situation. In the short answer test format, the test takers were required to listen to stimuli and questions and they did not have to read anything after that. Thus, we could conclude that the short answer test format, in this present study

situation, required less linguistic and nonlinguistic knowledge (memory and experience) from the test takers because they were required to listen only. To answer the short answer questions, the subjects did not have to produce any answer longer than two words and wrong spelling was accepted as long as their answers could be understood; therefore the task format required less test takers' knowledge. In summary, the MC format required the test takers to perform listening and reading tasks at the same time and the fact that this demanded more linguistic knowledge, memory and test takers' experience caused the multiple choice format to be more difficult than the short answer test format in the present study.

2. Viewing the questions and options required more information processing from the test takers. This can be explained from a cognitive processing point of view (Vandergrift, 2004). Listening comprehension is subject to the limitations of human memory capacity. In an EFL test taking situation, a large proportion of the available resources in working memory is taken up when the test takers struggle with language difficulties. Especially, beginner-level L2 listeners have limited language knowledge therefore little of what they hear can be automatically processed. They need to consciously focus on details of what they hear and this maximizes the limitations of working memory and the speed of speech. Thus, their comprehension suffers from this limitation of working memory (Vandergrift, 2004). With multiple choice test format, the test takers had to understand the message, questions and choices at the same time. The result was slowing down the processing which caused memory loss or left no time for processing meaning. While, with the short answer format, in this present situation, the test takers had more time to process because they did not have to focus on choices given but they could write what was heard immediately to answer the questions. This is an advantage of the short answer format over the multiple choice task in the present study of listening comprehension test.

In summary, the results from the analysis show that making different use of response format has a significant result on the scores of the test takers. The multiple

choice format, which is mostly favored by test writers and test takers, requires more knowledge and makes considerable processing demands on test takers. The multiple choice test format; therefore, is more difficult than the short answer response format in the listening comprehension test. However, this conclusion should be carefully undertaken because the difficulty level of each question in both formats was not strictly controlled as mentioned in the limitations of the study.

Question 2

“Do English accent varieties used in listening input have a significant effect on test takers’ listening comprehension ability? And if they do, how much is the effect size?”

Hypothesis 2

“There is a significant difference in the scores obtained from the test using English native speakers’ model as listening stimuli and the test using other accent varieties of English as listening stimuli.”

The results from the ANOVA analysis reveal a significant effect of using accent varieties of English as listening stimuli in the test. There was a significant main effect of test version and the test version A which used native-speaker voice was easier than the test version B that used nonnative-speaker voice for the test takers in this present study. The analysis suggests that accent varieties of English used as listening input differentially affects EFL listeners’ comprehension ability.

The findings coincide with Major et al.’s (2005) results. Major et al. (2005) also found out that ESL listeners experienced the least difficulty with the American English and greater difficulty with the international English dialects. In their study, ESL listeners performed notably better on tests with lectures delivered in American English than on tests with lectures delivered in international dialects. These results suggest that ESL and EFL listeners found international varieties more difficult to comprehend than native speakers’ varieties.

Although the scores of both versions were reported to be significantly different, the result produced a very small effect size of $d = 0.29$. A lot of test takers who could score high in their final examination test for mechanical matching showed a very little difference in their ability on test version B which was the nonnative speaker version. This means although it was more difficult to hear and pay attention to unfamiliar accent varieties, they did not obstruct comprehensibility of the message for the advanced test takers. This could well explain the result of a negligible effect

size value of $\eta^2 = 0.021$ for the difference between the scores from the two versions. If more advanced test takers were included in the study, the result might have been different. This is well supported by Munro and Derwing (1999) and Munro et al. (2006) that speech marked by nonnative segmentals and prosody appears to be understood by native listeners just as well as native produced speech from a familiar dialect and even an utterance that is rated as heavily accented might still be understood perfectly. All these findings imply that although using accent varieties as listening input could lead to the difference in the test takers' scores, the difference was very tiny. This difference in the scores might be eliminated if the test takers are provided more chance to become accustomed to these accent varieties. If efficient EFL listeners are more familiar to listening to these accents, there should be no difference in the scores for the comprehensibility of nonnative varieties of English.

Why was the test takers' average scores obtained from the native speaker test version different and higher than those from the nonnative speaker test version?

The average scores from the native speaker version were found to be different and higher than the average scores from the nonnative speaker version with a small effect size of the test version ($d = 0.29$). There are three possible reasons which could be stated to support the listening comprehension test situation in the present study. The three reasons are as follows:

1. The first reason for the difference in the test scores between the native speaker and nonnative speaker test version may have to do with language attitudes, which have potent effects on listening comprehension (Dalton-Puffer et al., 1997; Gill, 1994; Mackey & Finn, 1997; Pihko, 1997). In the previous studies, the researchers found that the more prestige accorded to a language or dialect, the better the comprehension of it. Munro et al. (2006) concluded that listeners might respond negatively to accented speech because of impatience, inexperience with nonnative speakers, or prejudice. In this present study, the attitudes of the test takers were measured through the semi-structure interviews. Apparently, the native varieties, especially American and British English, bear the most social prestige for the test takers in status dimension. American and British

English received the mean scores of 3.16 and 3.15 respectively, while none of the nonnative varieties received the mean score of more than 2.66. Thus, it is not surprising to see the test takers of the present study performed better in listening comprehension ability to listening stimuli delivered in native speakers' English version compared to the nonnative speakers' version.

2. The second reason has to do with familiarity (Flowerdew, 1994; Wilcox, 1988; Major et al., 2005; Munro et al., 2006). The test takers might experience difficulty in understanding speech that differs from the patterns of oral production to which they are accustomed to. The EFL test takers in this study have been exposed more to native varieties of English in classroom and media. They are presently university students in Thailand whose exposure to nonnative speakers' varieties outside the classroom has been limited. Although international varieties of English are widely spoken by nonnative speakers in Thailand, it is not commonly encountered in academic settings which are the test takers' environment. It is rare to find nonnative speakers besides Thai lecturers to teach English at a university level. Most of the English lecturers, besides the Thai teachers at universities, are native speakers; particularly, at UTCC where the main study was conducted, all foreign teachers who teach English are native speakers. Moreover, from the interview results which will be discussed later in this chapter, it is obvious that familiarity is the main reason for their preference for native accent varieties. More than half of the interviewees said they wanted to listen to only native accent varieties in the test because these varieties were what they were familiar with and exposed to in class.
3. In addition to familiarity, it is widely known that attitudes towards different accent varieties influence comprehension in listening. Positive attitudes aid comprehension, while negative attitudes interfere with comprehension (Major et al., 2005). Although successful L2 listening comprehension is not attributed exclusively to the existence of positive attitudes towards the English varieties, we cannot deny that it has some

influence at a certain level. In this present study, the test takers were reported to have more positive attitudes towards the native accent varieties than the nonnative accents, and they performed better in the test that used native speakers' voice. Some of the test takers mentioned their negative attitudes towards using nonnative speakers' voice in the test. They stated that they did not want to do the test when they heard the nonnative speakers' voice. However, they realized later that the speech could still be understood; unfortunately, they did not listen to the test carefully from the beginning. Obviously, their negative attitudes obstruct their true performance in listening comprehension. On the other hand, the students who listened to native speaker test version did not show any negative attitudes towards the voice in the test. This present study reveals some evidence for listening comprehension and listeners' attitudes. It seems that stereotypes regarding nonnative accented speech exist in the mind of the test takers, and these attitudes may have some influences on listening comprehension.

Question 3

“Is there a significant interaction effect between different listening test formats and varieties of English? And if there is, how much is the effect size?”

Hypothesis 3

“There is a significant interaction effect between the test formats and the varieties of English on listening comprehension scores.”

Findings regarding the interaction between test formats and test versions indicate that there is an interaction effect between the two main variables. The means were plotted to show the interaction pattern and it can be interpreted from the graph that the multiple choice format variable did not produce much difference in scores in both test versions, whereas the short answer format could make a bigger change in the scores especially when it was combined with native speakers' varieties in version A. This was confirmed by the results from the *t*-test (t value = .612, p = .541) where the multiple choice format produced similar scores in both test versions. Combining native or nonnative speaker varieties with the multiple choice format did not have a significant effect on the test takers' listening comprehension scores.

On the other hand, the mean scores for the short answer format with test versions A and B looked quite different and the results from the *t*-test (*t* value = 2.942, *p* < .01) confirmed their difference. The two means were significantly different which means that combining accent varieties with the short answer format had an effect on the scores. Apparently, the combination of the short answer format and native speaker voice stimuli could produce an easier form of listening comprehension test.

On the other side, considering the aspect of varieties, using native speaker voice input with multiple choice and short answer format seems to show a big difference when looking at the mean scores. The results, which are *t* value = 5.135, *p* < .01 demonstrate that the combination of native speaker varieties with different formats gave different results and short answer format provided higher scores. Another comparison to be considered is the combination of nonnative speaker varieties with the two formats. The means for SA format are again quite close, and the *t* value of 1.889 and *p* value of .062 reveal that, unlike native speaker voice input, using nonnative speaker voices as test stimuli gave similar results in scores in both test formats. That is for the test takers, nonnative speaker varieties were still difficult to comprehend and changing the test format did not give much assistance to them.

To conclude, the interaction effect result shows that the combination of the two variables made a difference in the test takers' scores. The test takers did not perform differently in the multiple choice test format in both test versions. On the contrary, they performed differently in the short answer format when it was combined with native speaker test version. Considering the use of accent varieties in the test, the test takers performed indifferently in the two formats when the test came with the nonnative speakers' voice.

Why was the average score of the test takers obtained from the combination of native speaker test version and the short answer format different and higher than those scores obtained from the combination of nonnative speaker test version and the multiple choice test format?

The interactions between test format and English varieties variables are in accordance with previously discussed findings. There are two main possible causes which can explain the pattern of interaction:

1. The multiple choice format was found to be demanding for memory capacity of the test takers because viewing the questions and options required more information processing demand. When the demanding task like the MC format was combined with nonnative varieties of English, this combination required even more cognitive processing from the test takers. This is because the test takers faced more difficulty in (1) understanding the speech sound with which they are not familiar, (2) understanding L2 with their limitation of L2 linguistic knowledge, and (3) understanding questions being asked and at the same time trying to make sense of options provided with their linguistic deficiency. Unsurprisingly, the combination of two demanding factors that are the MC test format and nonnative varieties of English produced the most difficult test version of listening comprehension in this present test situation.

2. Native varieties of English were reported to be more familiar and easier to comprehend when compared to nonnative varieties. The test takers were comfortable to listen to native speakers as this is very similar to what they usually hear in their classroom situation so they did not reject to the familiar speech sound. When this familiar speech sound was combined with a less demanding task like the short answer test format, the test situation was friendlier to these test takers. With the short answer format, they were required to write only a few words in which they were told to be less serious about spelling. Once they listened to a familiar native speakers' voice that talked about familiar topics, it is undoubted that they could perform the correct answer right away if they understood the questions. Consequently, the combination of the short answer format and the native varieties of English created a friendly test atmosphere that led to the production of the easier test version of listening comprehension.

Question 4

“What are the test takers’ attitudes towards accent varieties of English and their preferences towards using accent varieties of English as listening comprehension verbal input?”

Concerning listeners’ affective reactions to the varieties of English, attitudes and preferences towards these accents were investigated separately in the semi-structured interview. Although not all the test takers were included in the interview, the results obtained were systematic and meaningful. Since the answer for this research question is rich of information, the result could be categorized and discussed in five topics as follows:

a. Native varieties are prestige varieties

The responses showed that, on average, native varieties were rated as having more positive attitude than nonnative varieties. In status dimension, particularly, American, British and Australian English received much higher scores than those of Malaysian, Japanese, Singaporean and Chinese varieties. Taking a closer look at the adjectives describing ‘status’, the words –elegant, educated, intelligent, wealthy and successful - convey prestige of varieties. The findings concur with the classic pattern of such research that reflects the tradition that prestige accents score highly in power/status variables (Bayard et al., 2001). For Thai students in this present study, this means that a person who sounds like a native speaker would be perceived as having ‘prestige’. This claim is confirmed by some responses provided by the interviewees when asked about their preferences towards varieties. There were 13 respondents who mentioned directly that a person, who sounded like a native speaker looked smarter, more educated, and had a better personality than a person who did not sound like a native speaker. Definitely, native varieties of English are the prestige varieties for the respondents.

b. The American variety – the English variety of preference

The findings from the interview showed that American English is the most popular model for the respondents. The American variety received the highest scores in solidarity dimension and gained the first rank in both dimensions concerning attitude study. This was quite reliable since the subjects responded to these voices without being told the nationalities of the speakers. Nine respondents mentioned that

American English sounded friendly to them and was the easiest among other varieties to comprehend. Three interviewees said they liked the American accent, and another five named an American lecturer whose voice was preferable. The results were not unexpected, actually, they were of the same manner as the results of previous studies. In Bayard et al. (2001), the American accent was well replacing RP as the prestige and preferred variety, and this was happening in New Zealand, Australia and some non-English-speaking nations as well. It was stated that the American variety is the preferred and powerful variety in Sweden, Norway, Thailand and Singapore (Bayard et al, 2001).

The American dominance in all media guises such as fast food, pop music, films, TV sitcoms, and American ownership of mega multinational companies is the potent factor for the present study results. As Thailand is a monolingual society in which a majority of Thais speak their native language at home, in schools and in offices, the only place that Thai students are equipped to hear English outside their classrooms is from the media. The respondents are most accustomed to the American variety from their more frequent exposure to the American accent than other English varieties. Consequently, it is not surprising to see that they rate the accent with which they have become familiar at school and from media best.

c. The Japanese variety – difficult but pleasant

Another interesting point to discuss here concerns the Japanese accent. For the solidarity dimension, the Japanese voice was rated quite high at almost the same rank as the British accent (mean = 2.74 for British and 2.73 for Japanese). This might be because the respondents who are young Thais have a positive feeling towards things about Japan. This claim was supported by the interviewees' reflection when asked about their comprehension on different varieties. Approximately one third of respondents stated that though they did not understand much when listening to the accent variety (they did not know that it was Japanese), they found it was 'cute' and 'pleasant' to hear. To them, the accent did not sound 'confident' (stating power) but humble and friendly (stating solidarity). A few interviewees said they would be 'patient' to try to understand since the voice sounded so lovely. It was proposed by Munro et al. (2006) that interlocutors might respond negatively to accented speech because of impatience and inexperience with L2 speakers. Since young Thais are more accustomed to Japanese guises (in a lot of media, pop music, cartoons, products,

entertainment, etc) than to other three nonnative accent varieties in this present study, for this reason, they did not respond so negatively to Japanese-English utterance and rated the variety as high for the solidarity dimension.

d. Stereotype on varieties

The study revealed an apparent result that nonnative varieties of English are more difficult for the test takers to comprehend. Further, this is confirmed by the interview part, 80% of respondents inferred directly that nonnative varieties are harder to understand than the other three native English accent varieties.

Although the test takers could distinguish between native and nonnative English, not many of them could tell the difference between each English variety. They, especially, did not really know the difference between general American and British English. For instance, when they said they liked the British variety, instead they gave examples of American people and sometimes an Australian. This might be because of the fact they have been provided with a mixed exposure of native English varieties. At UTCC, a lot of commercial textbooks and teaching materials used are examples of mainly British and American English, and these materials are taught by a mixture of native English lecturers from the UK, the USA, Australia and a few from Canada, and New Zealand. However, the students have not been exposed to a particular variety long enough to tell the difference among native speakers, and they might not even pay attention to this matter. This might lead to their inability to name correct native varieties of English. A number of respondents mentioned that they loved to listen to British English and some of them even said they wanted to sound British but actually they just meant they wanted to be native-like. This is the evidence that native speakers of English were often stereotyped as British people for the respondents.

The discussion above reflects the respondents' fixed stereotype of the native speakers' utterance. There are still a number of students who are influenced by British imperialism. Although Thailand was not colonized by UK in the past, it cannot reject that once dominant power such as British English which had exercised control over a colonized area of Thailand's neighbors and a number of countries worldwide also has a strong impact on some Thai people's preference for British English. This claim is well-supported by a lot of respondents' opinions (eight of them), when asked the reason for their preference of British English, all of them said,

“British English is original because British people are the owners of the language. It is very prestigious to sound like an English person.” However, it is important to note here that this is their personal beliefs. In reality, as stated in the findings, most of the respondents did not really know how the British accent sounds like. This was revealed by the matched-guised technique that actually they preferred the American accent more than the British accent.

In terms of nonnative varieties, from the interview findings, it was revealed that a number of respondents had stereotypes about the accent varieties in that they found they were hard ‘to catch’. When asked whether they could recognize the varieties used in the test stimuli, about one third of them were certain that Indian English was included in the test (actually, there were no Indian speakers in this study). From the crosstabulation, the results showed that most of the test takers were unable to recognize the difference among accent varieties. Therefore, this means that Indian English accent was stereotyped as a hard-to-understand accent by the test takers. This rigid stereotype was mentioned again when the respondents were enquired to name any English varieties that they found difficult to comprehend. The Indian English variety was ranked in the top by six respondents. Perhaps, they might think that the voice examples from Malaysia were Indian English.

However, the participants, who have had experience English in native environments, responded to these questions a little differently. They seemed to be open-minded on varieties. All of them were more capable in distinguishing the difference between American and British accents, and none of them expressed special favor for British English or strong belief about Indian English. It is interesting to note that those who mentioned the British English as their model (without knowing how real British English sounds) have not been in an English-speaking country. This confirms the result from Dalton-Puffer et al.’s study in Austria (1997) that the evaluations of the students with EFL experience reflect rather rigid stereotype, while those students who have spent some time in English-speaking countries reveal more individualized, situation-linked attitudes.

e. Native speaker accents were preferred for listening comprehension tests

Because of their familiarity and positive attitudes towards native English varieties, two third of the respondents insisted that they wanted only native varieties to be included in the test. On the other hand, another 33 % of respondents thought

that a listening comprehension test should incorporate varieties both from native and nonnative speakers. They realize that in real world there are more nonnative speakers and they wanted to expand their comprehension ability to include these different English accents. Some of them mentioned their excitement and amusement when hearing varieties. It is obvious that those students who agreed with the idea of incorporating nonnative speakers' voice were more advanced students, and they did not care much about their performance on the test. On the contrary, the participants who scored lower were more worried about not being able to comprehend unfamiliar utterance which would affect their test scores. These lower-scores participants did not want to score low in the test, whereas those who were more advanced were more concerned in the test authenticity and wanted to be challenged by something new in the test.

Another interesting point to discuss is the fact that more than half of the respondents were not uncomfortable with accent varieties in a non-test situation. Some respondents, who disagreed with the idea of incorporating nonnative speakers' voice with the test, proposed they would love to have more practice in listening to nonnative varieties in classroom. One respondent stated clearly that she thought if she had been prepared to listen to nonnative varieties more in class, she could have scored higher in test version B. This leads to the idea that nonnative varieties will be more acceptable if the test takers are provided with more chance of exposure. Familiarity, as stated previously, still plays a significant role in the test takers' attitudes towards using varieties of English as listening comprehension test stimuli.

III. Recommendations

a. For Research Consumers

The results obtained have direct implications to the construct validity of listening comprehension tests. In regards to the results of the present study, the following points for listening comprehension test construction are worth being noted.

1. The practical question underlying this study is whether it would be desirable to use other varieties besides the native speakers' varieties in listening comprehension tests. The results of the present study, which shows the effects of English varieties on listening comprehension, have implications for

teaching and testing English listening comprehension for L2 learners, to be more specific- EFL learners. Worldwide, most listening comprehension tests are limited to the Standard American English variety or the British English variety. This study has addressed the problem of bias against foreign-accented speech. It shows that bias is a factor in some of the listeners' judgments and it could have a meaningful effect on listeners' comprehension ability. The fact that there are many more nonnative speakers of English than native speakers around the world, and this tremendous number has led to a focus on English as an International Language (EIL). Therefore, EFL students should be allowed to be exposed to nonnative speakers' varieties of English more in class to enhance their familiarity with these varieties. There are some research findings which show that exposure to accents can increase comprehension of said accents and listeners might benefit from explicit training on accented speech, especially in terms of attitude towards them (Gass and Varonis, 1984; Derwing et al, 2002). Given this, it would seem reasonable for an achievement listening comprehension test to include other varieties as listening stimuli as long as students are provided enough exposure to them. Furthermore, because an achievement test should test what is taught in class, *it is recommended that EFL listening teachers should provide more opportunity for their students to be exposed to varieties of English.* If students are provided with enough exposure to other varieties, it would be valid and appropriate for a test to include other varieties of English other than those native speakers' only. Including varieties of English would result in a more authentic and valid listening comprehension test.

2. The practical goal of determining whether changes in test formats would result in differences in test performance was set out. In most situations, especially with a high-stakes test, the multiple choice format is a favorable form of test. The results of the study reveal that the multiple choice method might pose threats to the construct validity of the test if the difficulty of the format is uncontrolled. The findings suggest that the less able test takers might be put at a disadvantage because of the choices provided. On the other hand, with the short answer method, these students do not have to worry about understanding choices while listening. Therefore, the short answer format allows them to

perform more to their true comprehension ability. Further, it is reported by Yi'an (1998) that the multiple choice method allows much uniformed guessing. Often, test takers give correct answers for the wrong reasons, so it does not guarantee test validity. To balance the disadvantage that might occur from using a solely multiple choice format, the short answer format might be an alternative particularly for a low stakes test situation like a classroom activity or a quiz because teachers would need some time to do the marking. In conclusion, *it is recommended for L2 listening teachers that listening comprehension tests should present a variety of test formats to balance and avoid bias towards individuals.*

3. The core of the concern for the test takers of listening comprehension test is the fact that comprehension involves much more than just the application of linguistic knowledge. For the listening test writers, perhaps they need to clarify exactly what they want to test. Whether the test task format used really measures the test takers' ability of listening comprehension and whether the test format used is suitable and practical for the test situation purpose should always be the concern. The short answer format, which is usually more difficult to administer with a large number of test takers, is recommended to be used in class quizzes or in test situations where teachers have more time for marking with fewer students. However, *it is strongly recommended that the test task itself should not demand the test takers' memory capacity with long text and should not require long answer, unless that particular listening test has different objectives of testing.* Given this, the temptation is to suggest that listening tests should concentrate on measuring the test takers' ability to understand the propositional content of text, not their background knowledge, reading ability, or writing ability.

4. The authority to create Standard English no longer resides solely with its native speakers. Lowenberg (1993) indicated more than 10 years ago that there were potential cases where the control of the future use of English would be passed to nonnative speakers. This is indeed the case in many areas of the English-speaking world. Jenkins (2006) called for a major rethinking of English language goals and the overhaul of English language testing due to the

recent changes in both users and uses of English around the world. There is an attempt in conducting the standard framework for English for international uses, for instance, as seen in a work done by Nakano (2007) and many other more. Language testers are facing the changing face of English. However, the stakeholders of tests can be fairly conservative as it can be seen from the present study results and other studies. Jenkins (2005) found out that teachers from the expanding circle did not really wish to use their accented English to express their L1 identity. Nemtchinova (2005) also reported that ESL students in the USA prefer to be taught by native speakers. This is concordant with the present study findings that the EFL students or the test takers still have strong belief in the norm of native speakers. This might gradually change, but not suddenly, because the people's past experience in classroom teaching, media, and social association combines to affect their attitudes to English at the deeper level. As it is noted by Taylor (2006) that over the next 10 or 20 years, emerging Englishes including EIL may well grow in status and take on a role as pedagogic and assessment models for English learners. While we are waiting for this change, *it is recommended for English teachers to nurture listeners by increasing familiarity and appreciation for world Englishes because it is obvious from the present study results that familiarity and appreciation of varieties could lead to better listening comprehension.* In the listening classroom, teachers can provide resources to create a greater understanding of the concept of international English among students and this can then be passed on to the listening test situation.

b. For Further Studies

A number of areas that could be executed in future studies are given below.

1. This present study concentrates on the listening test format methods on paper work assessment. The test takers were allowed to listen to only stimuli and questions but they did not have any hint from visual elements. It is interesting to investigate whether adding visual elements would have any effect on listeners' comprehension ability when compared to the situation in which test takers are given no support of visual input. An expansion of research into the impact on listening comprehension of visual

element in both using pictures or using video in the electronic media would be useful for a more valid construct of a listening test.

2. The refocusing on the learner and on learning has created an explosion of research aimed at investigating learners' characteristics. As the present study focus on only two variables concerning test method and test input, listener factor variables, which could have an important effect on test takers' scores, is left open to be studied further. Recent advances are encouraging. Language assessment researchers have just begun to investigate the interaction between mental processing and second language test performance (Purpura, 1999). Studies on the relationship between individual differences such as test takers' cognitive styles and listening test performance are new to the field of listening ability assessment.
3. In regards to using varieties on English in the test, a foreign accent does not necessarily cause L2 speech to be low in comprehensibility or intelligibility (Munro and Derwing, 1999). It has been proven by the present study that it does have an effect on listeners' comprehension with a very small effect size. Given this finding, it makes sense to investigate what causes the differences in understanding varieties besides familiarity. There are a lot of factors that might play important roles such as listeners' background, bias, attitudes, language proficiency, environment, instruction, etc.
4. Concerning the attitudes towards varieties of English, it is shown from the present study interview findings that a lot of respondents believe that an Indian variety of English is difficult to comprehend. However, there is no empirical evidence to prove this strong belief among Thai people. It might be interesting to include the Indian English variety in future research to see whether it is really harder to comprehend when compared to other international varieties of English.
5. The present study provides some evidence of the effect of world media based on American models. Bayard et al. (2001) mentioned that American

English not British English will remain the major global form of English into the indefinite future. This has proven true for accent attitudes from the present study result. Whether the positive attitudes towards the American variety will add comprehension support to test takers when performing listening tests is left unanswered for Thai language learners.

6. The present study did not control strictly over the difficulty of the questions used in the two formats. Furthermore, the selection of native speakers' voice used was not firmly controlled over the background of these chosen speakers. Besides, due to the matched-pair technique, the researcher did not randomly assign the subjects to groups. It is suggested that the mentioned limitations be solved when doing a replicated study in the future.