# *The effect of including non-native accents in English listening tests for young learners: psychometric and learner perspectives*

David Wei Dai

ORCID ID: https://orcid.org/0000-0002-3575-131X

Submitted in partial fulfilment of the degree of

Masters of Applied Linguistics

School of Languages and Linguistics

the University of Melbourne

June 2015

# *The effect of including non-native accents in English listening tests for young learners: psychometric and learner perspectives*

David Wei Dai

## Abstract

As English has been used widely as a lingua franca for communication, language testers have started to evaluate the proposal for introducing non-native accents into the listening input of English tests. This study aims to further this debate from both the psychometric and learner perspectives by not only investigating how accents influence test takers' performance, but also eliciting their subjective perception of accents. 80 young L1-Mandarin test takers were recruited and divided into four groups, with each group listening to one accented version of the same test. The four accents used in this study were Australian, Spanish, Vietnamese and Mandarin English accents. Test takers subsequently completed a Likert-scale questionnaire, which measured their accent perception on three sub-scales, Familiarity, Comprehension and Attitude. Results indicate that the Mandarin accent group performed significantly better than the other three groups in the test and also perceived the Mandarin accent significantly more comprehensible, lending support for the shared-L1 effect. No significant difference is observed among the three non-Mandarin groups whether in the test scores or the Comprehension sub-scale. There is no significant difference in test takers' perception of the four accents in terms of Familiarity or Attitude. The central implication from this study is that there is potential for the inclusion of non-native accents into listening tests provided the shared-L1 effect can be properly addressed.

## Declaration

This minor thesis contains only original work by the writer, except for the references that have been appropriately acknowledged. No section of this thesis has been presented at conferences or appeared in any publications.

The length of this thesis, exclusive of tables, references and appendices, is 10,954 words.

_____

David Wei Dai

(Student ID: 632159)

## Acknowledgments

First, I would like to take this opportunity to express my deepest gratitude to my thesis supervisor, Assoc. Professor Carsten Roever, without whose invaluable guidance, unreserved support and limitless patience, I would not have been able to finish this thesis.

I would also like to thank Dr Averil Grieve for the excellent and constructive advice she provided, which helped me tremendously in thesis writing.

I am deeply indebted to the 12 accented speakers and 35 listeners in this study who took the time and effort to help me voluntarily.

I am also grateful to the 80 students who participated in this study, the students' parents who supported this study, the coordinating English teachers and the school principal who helped me to conduct this study.

Finally, I would like to express my thanks to my parents and family members who have loved me unconditionally and have always been there for me no matter what.

# Table of Contents

## List of Tables

## List of Figures

# Chapter 1. Introduction and Literature review

## 1.1 Introduction: English as a lingua franca and the need for including L2-English varieties into English tests

As English is being used globally in multicultural contexts, there has been a growing number of English as a Lingua Franca (ELF) researchers arguing that English tests should start assessing multidialectal competence instead of using several dominant English varieties as their benchmarks (Brown, 2014; Canagarajah, 2006; Canagarajah, 2007; Clyne & Sharifian, 2008; Jenkins, 2006; Jenkins & Leung, 2013; Seidlhofer, 2011; Sharifian, 2013; Taylor, 2006; Taylor & Geranpayeh, 2011). Such an argument is based on the observation that in the context of postmodern globalisation, English is a shared language among not only L1-English speakers but also overwhelmingly L2-English speakers (Seargeant, 2012). Since the number of L2-English speakers far exceeds L1-English speakers, language testers feel the need to investigate the possibilities of including L2-English varieties into high-stakes English tests for better construct representation.

In order to better represent the ELF reality in international English tests, the assessment paradigms of English tests need to be revised. In the case of listening tests, much research has focused on assessing the feasibility of introducing dialectal or L2 accents[1] into listening constructs that traditionally

[1] The dichotomy of native/non-native, L1/L2 English accents can be problematic as it overlooks cases like bilingual speakers. Nevertheless, the researcher considers the impact of bilingual test takers on large-scale international English listening tests to be fairly limited. Therefore, for the

only use few dominant native accents (Harding, 2012; Major et al., 2002; Major et al., 2005). However, such attempts have proven difficult as can be demonstrated by the fact that major high-stakes tests still limit their accent input to few standard native varieties (Cambridge English, 2015; ETS, 2015). However this practice has been criticised repeatedly by language testers as they cannot foresee a domain where test takers only need to listen to such few accents (Harding, 2012).

Built on previous research, this study aims to examine the practice of introducing non-native accents into listening tests for young learners. In order to triangulate the research findings and offer a more comprehensive view, this study utilises two main research instruments, an objectively scored listening test from the psychometric perspective and a Likert-scale questionnaire from the learner perspective. The test looks into if four groups of L1-homogenous test takers will perform differently in four accented versions of the same test while the questionnaire investigates if test takers will perceive the four accents differently in terms of accent familiarity, comprehension and attitude. The connection between the psychometric and learner perspectives will be explored as well. Findings from this study are expected to offer insight for multidialectal listening assessment in language testing, teaching and decision-making processes.

---

sake of simplification, this study still adopts such dichotomous classifications. To be more specific, native/L1 English accents refer to the accents of speakers who use English as their first language while non-native/L2 English accents refer to the accents of speakers who have learnt to speak English after their first languages have already been established and whose English accents carry phonological features of their L1s.

### 1.2 The psychometric perspective: research related to the shared-L1 effect

From the psychometric perspective, one of the challenges of introducing L2 accents is the concern of a 'shared-L1 effect', a language testing term formally defined in Harding (2012, p. 164) as a phenomenon where a certain group of test takers tend to comprehend the listening materials more easily and consequently perform better in a test when they share the same L1 with the speaker of the test recording. If such an effect holds true, it will be considered test bias as a listening test construct should only measure test takers' listening competence and test takers should not be advantaged because of their first language backgrounds.

Though the shared-L1 effect is largely a language testing term, the relationship between accent and comprehension has actually been extensively investigated in different forms in fields such as psychology (Sidaras, Alexander, & Nygaard, 2009; Stevenagea, Clarkea & McNeillb 2012) and phonetics (Stibbard & Lee 2006; Weber, Broersma & Aoyagi, 2011) aside from language testing (Harding, 2012; Major et al., 2002; Major et al. 2005; Ockey & French, 2014). While the cause for the shared-L1 effect is still inconclusive, research in psychology has pointed out a potential explanation when a wealth of studies noticed that familiarity with an accent can lead to better comprehension and recognition due to repeated exposure (Adank et al., 2009; Bent & Bradlow, 2008; Sidaras, Alexander, & Nygaard, 2009; Stevenage, Clarke & McNeill, 2012; Weber, Broersma & Aoyagi, 2011). This familiarity advantage can partly explain the shared-L1 effect because test takers of a particular L1 background are generally assumed to have more access to their L1 English accent than other accents.

It is still contentious as to what extent this shared-L1 effect can actually cause noticeable difference among different test taker groups. Findings from

some studies show that only certain L1 groups will conspicuously benefit from a shared-L1 listening recording and this L1-advantage is not generalisable to all L1 groups (Harding, 2011; Major et al., 2002). However, other researchers have found little difference in terms of comprehensibility whether or not the recording is delivered in listeners' L1 accent (Abeywickrama, 2013; Butler, 2007; Munro, Dewing & Morton, 2006; Nejjari, Gerritsen, Van der Haagen & Korzilius, 2012).

In terms of assessing multidialectal listening competence in the testing field, there are two recent studies that are of particular pertinence. The first study conducted by Harding (2011) offered mixed findings to the shared-L1 advantage argument. Harding recruited 212 test takers (70 Mandarin-L1 participants and 60 Japanese-L1 participants) and recorded a listening test with one section in Australian English accent, one section in Mandarin English accent and the third section in Japanese English accent. After administering the same test to all of his test takers and using DIF to analyse individual item performances in the test results, Harding noticed that Mandarin-L1 test takers were conspicuously advantaged on several items with the Mandarin accent, lending support for a shared-L1 advantage. However, such a clear advantage was not observed in the Japanese-L1 test taker group. The mixed findings from Harding (2011) echo the results from Major et al. (2002). In Major et al. (2002), a Spanish-L1 test taker group benefited from the Spanish English accent, which supported the shared-L1 effect argument. However, a Chinese-L1 group in their

study actually performed worse with the Chinese[2] L1 accent compared to other test taker groups, which contradicted the shared-L1 advantage.

The second study by Ockey and French (2014) is based on a large-scale research project in multidialectal[3] listening research. They recorded one TOEFL common lecture in nine different accents (one US, four Australian and four British English accent speakers of varying accent strength) and administered the test on 21,726 TOEFL test takers. Using ANCOVA to investigate the interaction between accent strength and accent familiarity with test scores, they noticed a general pattern where test takers' performance decreased when accent strength increased or when accent familiarity decreased. It should be noted that instead of grouping test takers based on their L1 backgrounds, Ockey and French divided test takers based on their self-reports as to how familiar they were with the three accents (US, Australian and British) in question.

In summary, the studies in multidialectal listening testing so far have generally relied on listening tests as objective research instruments to investigate how accents influence test takers' performance.

### 1.3 The learner perspective: research related to accent perception

If non-native accents should be introduced into high-stakes English tests, this practice would cause multifaceted impact on language testing, teaching and learning. **Section 1.2** aimed to offer an account of existing studies that

[2] In Major et al. (2002) they used the term "Chinese" while in Harding (2012) he used the term "Mandarin Chinese". It is not clear from either paper how the Chinese accent was determined as Chinese accent is not a homogenous concept due to the vast number of Chinese dialects. In my study I use the term "Mandarin accent" specifically to refer to speakers who are proficient in Mandarin Chinese and were born and raised in mainland China.

[3] In this paper "multidialectal" is used in a general sense as encompassing standard native English accents, regional English accents, ethnic English accents, non-native English accents, etc. However my study has a particular focus on the introduction of non-native English accents.

investigated multidialectal listening assessment from an etic, objective, psychometric, language-tester's perspective. However, there is a lack of research into this topic from an emic, subjective, learner's perspective. Such a perspective would be interesting to look at for three reasons. Firstly, the issues unresolved in the etic perspective, such as the shared-L1 effect, might find their answers in test takers' subjective reflection. Secondly, the emic perspective empowers test takers, who are important stakeholders in the language testing industry and whose opinions on multidialectal assessment should be taken into consideration by test makers. Third, test takers are directly involved in any changes in language testing, teaching and learning and their perception of non-native accents can influence how the whole system operates. Therefore, aside from trying to evaluate the practice of introducing non-native accents into language tests from a testing perspective, it is of equal importance to examine this practice based on test takers' perception of accented listening tests.

Though there is a paucity of research in multidialectal listening assessment from learner's perspective, there are studies that investigated listeners' responses to accented speech samples, mostly under the topic of accent attitude. Numerous studies have noted that listeners generally show a preference for native accents of a certain language (Fraser & Kelly, 2012; Hendriks, Meurs & Groot, 2015; Hendriks, Meurs & Meji, 2015; Hiraga, 2005; Kim, 2007; McKenzie, 2008; Nejjari, Gerritsen, Van der Haagen & Korzilius, 2012; Zhang, 2009). In the specific context of English language teaching and testing, learner's favourable attitude towards the inner-circle English accent varieties has also been documented (Butler, 2007; Chien, 2014; Yook & Lindemann, 2013). However, it is worth investigating how such a preference is established.

Existing research has found that accent attitude is a complex issue often interrelated with factors such as accent comprehension, identification and familiarity.

In order to investigate the relationship between comprehension and attitude, Butler (2007) recruited 312 elementary-school Korean-L1 students and grouped them to listen to oral materials recorded in either Korean English accent or American English accent. A comprehension test based on the oral materials was then administered on the students. After the comprehension test, students were asked to report their attitude to the two accents in an attitudinal questionnaire. Results from the comprehension test indicated that the students showed no difference in their comprehension of the oral materials whether they were delivered in Korean or American accents. However, the group listening to the American accent responded more favourably in the attitudinal test than the Korean accent group.

Chien (2014) is another study that is of particular relevance to this debate. Having recorded speech samples in Australian English (AE) accent, General American English (GAE) accent, Indian English (IE) accent, Japanese English (JE) accent, Spanish English (SE) accent, Standard Southern British English (SSBE) accent and Taiwanese English (TE) accent, he found that his 317 Taiwanese-L1 listeners reported an overwhelming preference for the GAE accent but not to other native accents such as the AE and SSBE accents. Chien's explanation for such strong favouritism towards a particular native variety is that in the context of English teaching in Taiwan, there is a strong association of prestige and status with the GAE accent. It is worth asking if pedagogical and societal preference for certain may affect listener's partiality towards them as

well, which may have little to do with whether such accents are native or non-native.

Other studies in the field of accent perception showed that the impact of listeners' subjective responses to accents could be so influential that listeners' idealised perception of native accents could lead to unfounded, biased interpretations of other non-native accents (Hu & Lindemann, 2009). Such a phenomenon can even happen with professionally trained raters of speaking tests as reported by Winke and Gass (2013). They found that oral raters' objectivity could be affected by their identification of test takers' L1s from their accents, which implies that accent attitude is also related to accent identification. In terms of research specifically focusing on accent identification, McKenzie (2015) showed that listeners tended to first identify if the speech sample was from a native or non-native speaker based on phonological features before moving to more detailed classifications. Another study from Atagi and Bent (2013) reported that non-native English listeners were less accurate with accent identification compared to native English listeners but non-native listeners did demonstrate a heightened sensitivity to their L1 accents, which can be explained by the shared-L1 effect phenomenon discussed above. As to the relationship between familiarity and attitude, Scales, Wennerstrom, Richard and Wu (2006) reported a near perfect positive correlation between the most comprehensible accent and the most preferred accent.

### 1.4 Gaps in previous research

Based on the studies reviewed in **Section 1.2** and **1.3**, here I (the researcher) propose four gaps unaddressed in previous research.

**Gap 1**

Studies in the testing field have predominantly recruited adult test takers (Harding, 2011; Major et al. 2002; Major et al. 2005; Ockey & French, 2014) while to the best of the researcher's knowledge, there has been no studies on how young test takers respond to multidialectal listening tests. However, assessment of young learners is becoming an increasingly active field of research and practical work (Papageorgiou, Xi, Morgan & So, 2015).

**Gap 2**

To the best of the researcher's knowledge, there have been no studies that addressed multidialectal listening assessment both from the etic and emic perspectives.

**Gap 3**

There is a lack of research instruments that investigated global accent effects with sufficient items in multidialectal listening assessments. A limited number of items means that researchers will have difficulty computing reliability and may lead to a stronger content effect. A number of accented listening tests looking into global accent effects have been conducted but they were limited by the item number of their instruments. In the study by Ockey and French (2014) only 6 items were used and similarly Major et al. (2002; 2005) used 4 items per accent. In contrast, Harding (2011) utilised 30-40 items for each accented section in his test but chose to analyse test takers' performance at the item level with DIF instead of the global level.

**Gap 4**

With reference to the research instrument for investigating accent perception, most previous studies have used stimuli materials to elicit listeners' responses.

The speech materials in both Chien (2014) and Yook and Lindemann (2013) were 20-second pre-existing samples from the Speech Accent Archive collected by George Mason University (Weinberger, 2011). Though such samples can be authentic and easy to operationalise, the responses they elicited from listeners were based on momentary impressions of accents. In contrast, if listeners were provided with the opportunity to engage deeply and proactively with extended accented recordings, such as a listening test, it is expected that their accent perception could be more accurate and reliable.

### 1.5 This study: overview and rationale

This study aims to further the debate in multidialectal listening assessment and address the four gaps identified above. It recruited 80 15-year-old test takers of the same L1 background (Mandarin Chinese). The selection of young test takers in this study was intended to address *Gap 1*. In order to investigate the global accent effect in a listening test, the researcher prepared a 10-minute listening test and had it recorded in four different English accents, Australian English accent, Spanish English accent, Vietnamese English accent and Mandarin English accent. The test comprised 30 items in total, which aimed to address *Gap 3*.

The 80 test takers were divided into four groups with each group listening to one of the four accented test recordings. In other words, the test material was the same for four groups but the accent of the recording differed for each group. When the tests were completed, the four groups were asked to report their perceptions of their respective accented recordings in a 15-item Likert-scale questionnaire. Test takers' perception of accent in this study was defined as consisting of three constructs, their self-perceived familiarity of an

accent (Familiarity), their self-evaluated comprehension of that accent (Comprehension) and their attitude towards that accent (Attitude). Therefore the questionnaire consisted of three sub-scales with 5 items on Familiarity, 5 items on Comprehension and 5 items on Attitude. The combination of a listening test with a questionnaire served to address *Gap 2* and the extended test recording could effectively target *Gap 4*.

Another interesting point worth mentioning about the design of this study is the selection of accents. As this study aims to investigate the shared-L1 effect for an L1-Mandarin-Chinese group, it is self-evident that the Mandarin English accent was selected. As to native accents, Australian accent was chosen because the researcher wanted to focus on a less-dominant native accent so as to potentially offer different findings from previous studies that usually utilised British or American accents (Major et al., 2005; Ockey & French, 2014). The two non-native non-L1 accents were selected based on their phonological features and language families. As Mandarin is a Sino-Tibetan tonal language, the researcher wanted the two non-native languages to be non-Sino-Tibetan with one being tonal and the other one non-tonal so as to make the findings as generalisable as possible. Based on such a design, Spanish, a non-tonal Romance language and Vietnamese, a tonal Austroasiatic language, were chosen.

### 1.6 Research question

The purpose of this study is to offer some insight into the practice of assessing multidialectal English listening competence and as explained above, the researcher intended to approach this topic from both the etic and emic

perspectives. Therefore four research questions have been proposed for this study.

***RQ1****: What is the difference in the performance of four L1-homogenous young test taker groups in a listening test recorded in four different English accent versions?*

***RQ2****: What is the difference in how the four groups perceive the four accents in a questionnaire on Familiarity, Comprehension and Attitude?*

***RQ3****: What is the relationship between the objective listening test measurement and the subjective questionnaire measurement?*

***RQ4****: What is the relationship among the three sub-scales in the questionnaire?*

The connections among the four research questions are that RQ1 approaches the topic of this study from an etic perspective while RQ2 approaches the topic from an emic perspective. RQ3 aims to investigate the relationship between the etic and emic perspectives while RQ4 intends to analyse the internal connections among the measurements in RQ2.

# Chapter 2. Methodology

## 2.1 Pilot study

Prior to this study, the researcher and two of his colleagues conducted a smaller-scale pilot (Dai, March, Victor, 2014) on a sample from the same test taker population. In the pilot study, three accents were selected (British English accent, Mandarin English accent and Singaporean English accent) and a 20-item listening test was administered to three groups of test takers, each of which consisted of 20 students. A one-way ANOVA was used for data analysis with the test scores as the dependent variable and accent as the factor. ANOVA revealed the between-group difference was significant ($F_{(2,57)} = 6.221$, $p = 0.004$, $\eta^2 = 0.179$) and descriptive statistics are reported in ***Table 1***.

| Group | N | Mean | SD |
|---|---|---|---|
| **Mandarin** | 20 | 11.65 | 3.56 |
| **British** | 20 | 11.10 | 4.05 |
| **Singaporean** | 20 | 7.85 | 3.41 |

*Table 1: Descriptive statistics for the pilot study*

Post hoc test showed that there was significant difference in test scores between the Mandarin and Singaporean groups (p=0.008) and between the British and Singaporean groups (p=0.026). However, the difference between Mandarin and British was not significant (p=0.895).

The results from this pilot did not show a noticeable shared-L1 effect as both the Mandarin and British groups performed better than the Singaporean group. In retrospect, there were a few flaws in its research design. First, there

was a strong floor effect from test scores, which might lead to an underestimation of the effect size. Second, the selection of speakers was arbitrary and the recording of test materials was uncontrolled. Both the coordinating teachers and the researchers noticed that the Singaporean accent was stronger than the other two and the Singaporean speaker also spoke faster during the recording. Therefore the researcher suspected that because the Singaporean accent version was vastly more difficult than the other two, no difference between the Mandarin and the British groups could be observed. The flaws in the pilot study were addressed in this current study.

## 2.2 Instruments

### 2.2.1 The Strength of Accent Scale

The pilot study revealed that it is important to ensure that the strength of accent is consistent across all the speakers for the accented listening test. The impact of accent on test scores would be unreliable if one accent speaker, such as a Singaporean speaker, spoke with a heavy Singaporean English accent while another speaker, such as a Mandarin speaker, spoke with a light Mandarin English accent. In such cases, the researcher cannot determine if the difference in test scores was caused by the nature of accent (Singaporean accent versus Mandarin accent) or the strength of accent (a heavy accent versus a light accent). Therefore, an instrument needed to be developed to measure the strength of accent of speakers to ensure that all the four speakers for this study were of similar accent strength.

The Strength of Accent Scale (abbreviated as the Scale) was designed for such a purpose. The Scale was based on the accent scale used in Ockey and

French (2014), which relied on listeners' subjective judgment to measure accent strength. The Scale included five bands, with band 1 indicating listeners found the speaker having a very light accent and band 5 indicating listeners found the speaker having a very strong accent. When using Ockey and French's scale for piloting, the researcher noticed that the negatively phrased band descriptors in their scale tended to be confusing for listeners. Therefore, all the descriptors in the researcher's Scale were positively stated. The descriptors were also modified to include gradable adjectives so that listeners could have a clearer understanding of the nuanced differences between different bands. The Scale is included in Appendix A.

### 2.2.2 The Accent Strength and Identification Task

Aside from ensuring that all the four accented speakers had a similar degree of accent strength, the researcher also wanted to ensure that their accents were truly representative of their first language backgrounds. For example, it would be pointless to recruit a speaker for the Spanish accent when most listeners of this speaker's recording could not identify that speaker's English accent as a Spanish one. Therefore, the researcher designed an Accent Strength and Identification Task (abbreviated as the Task) to not only measure speakers' accent strength but also their accent identifiability. The listeners were required to take the Task and rate potential speakers' speech samples on accent strength and identifiability, based on which four final speakers would be selected for recording the accented listening test.

The Task comprised two sections, a Background Section and an Accent Judgment Section. In the Background Section, listeners needed to report their

self-perceived English listening skills, ability to differentiate different English accents and their level of familiarity with the four English accents in this study. Information about listeners' first language background was also elicited. In the Accent Judgment Section, listeners needed to rate the strength of speakers' accent based on the Scale provided. Listeners were given four accent identification options (Australian, Spanish, Vietnamese, Mandarin) and they needed to select an option that they thought best matched the speaker's accent. The Task is included in Appendix B.

### 2.2.3 The Accented English Listening Test

The main research instrument for this study was the 30-item Accented English Listening Test (abbreviated as the Test and included in Appendix C). The specifications of the Test targeted three key components in listening competence, as is displayed in ***Table 2***.

| Construct | Specifications | Requirements | Structure | Item Number |
|---|---|---|---|---|
| **1** | Understand the general information of a statement | Listen and choose the right picture | There are five sentences corresponding to five items | 1-5, 16-20 |
| **2** | Recognise main points in a long monologue | Listen to the passage and tell whether the following statements are true or false | There is one monologue for item 6-10 and another monologue for item 21-25 | 6-10, 21-25 |
| **3** | Locate and identify specific information in a long monologue | Listen to the passage and complete the following sentences | There is one monologue for item 11-15 and another monologue for item 26-30 | 11-15, 26-30 |

*Table 2: Test specifications for the Test*

As shown by the item numbers, the sections in the Test followed a design that measured test takers' listening skills in two cycles of Construct 1, Construct 2 and Construct 3. Both the researcher and coordinating English teachers agreed that the difficulty of the tasks increased from Construct 1 to 3 and therefore the sequence for the Constructs should be from 1 to 3 in order to ease students into tackling more complex listening sections. The reason why the researcher did not choose to carry out 1-1-2-2-3-3 with the Constructs was due to concern that measuring the same construct repeatedly might lead to fatigue and impatience for the test takers.

In terms of the designing of the Test, the researcher used materials that the students were familiar with as benchmarks. The selected listening texts and corresponding items were approved by the coordinating English teachers as appropriate in terms of item difficulty, format familiarity and background knowledge. In other words, the Test was formulated to be test-taker friendly to avoid construct-irrelevant variances that could be introduced by confronting test takers with unfamiliar task types. The Test was piloted on five students of the same year level and they reported that it was suitable for their ability. It was expected that an average test taker in this 80-student sample would score 20 out of 30 (each item counted as one point). The reason the researcher was especially cautious was to ensure that the Test would not be too difficult for test takers as the pilot showed that test takers at a low proficiency level were highly vulnerable to foreign English accents. In order to avoid the strong floor effect evident in the pilot, this time the researcher deliberately aimed for a score of

20/30 so that the expected between-group differences could manifest within a better score range.

#### 2.2.4 The Accent Perception Questionnaire

In order to elicit test takers' subjective perception of the accented tests from an emic perspective, the researcher designed The Accent Perception Questionnaire (abbreviated as the Questionnaire). The three sub-scales in the Questionnaire were accent Familiarity (item 1, 4, 7,10,13), accent Comprehension (item 2, 5, 8, 11, 14) and accent attitude (item 3, 6, 9, 12, 15). Items in the questionnaire were organised in cycles of Familiarity (item 1), Comprehension (item 2), Attitude (item 3), Familiarity (item 4), etc. Such a design was chosen to minimise the possibility that test takers might recognise the intended constructs the Questionnaire was to measure.

The Questionnaire was proofread by coordinating English teachers and revised several times after piloting on potential test takers to ensure that its language was readily understandable for the test takers. The Questionnaire was composed in Chinese, the test takers' first language, so as to ensure optimal message transfer. An English version is included in Appendix D.

### 2.3 Participants

#### 2.3.1 Listeners

35 listeners (18 native English speakers and 17 non-native English speakers) were recruited to take the Accent Strength and Identification Task. At the time of this study, all listeners were enrolled in bachelor's or master's programs in Australian universities, reported high familiarity with the four accents in this study, demonstrated advanced English listening skills, and showed interest in

accent judgment and identification. It is crucial that the listeners needed to be familiar with the four accents because if not, their identification of the accents would be highly unreliable.

### 2.3.2 Speakers

12 speakers (three per accent) were selected for the first round of speech sample recordings. All speakers were male and aged between 30-40. The reason for selecting speakers of the same gender and age group is to ensure that listeners and test takers should not be influenced by any variables other than speakers' accents. When recruiting the 12 potential speakers, the researcher relied on his experience of the four accents and selected speakers who possessed mild-to-average accents. The researcher also investigated the language profiles of the 12 speakers to ensure that their accents were truly representative of their own groups. The selection criteria were that the speakers should have resided in their L1 countries since they were born untill they finished tertiary education. In other words, they were users of their L1 for at least 22 years since infancy, which should lead to adequate L1 accent transfer when they spoke English.

All 12 speakers were asked to record a 20-second speech sample, which were subsequently rated by the 35 listeners in the Task. Based on the results from the Task, four final speakers were selected for recording the Test. Details of these steps will be explained in **Section 2.4**.

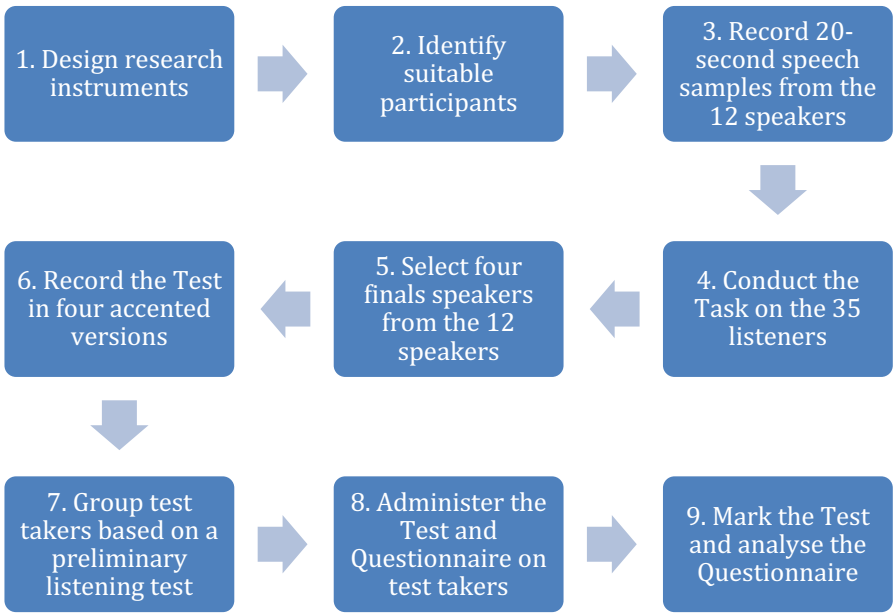
### 2.3.4 Test takers

All 80 test takers selected for this study were from a public middle school in a large city in mainland China. The 80 test takers (45 boys and 35 girls) were 15-year old English learners of Mandarin-L1 background. The coordinating English

teachers from that school assured the researcher that the 80 test takers were highly homogenous in terms of their English listening competence because they had similar learner profiles and studied the same English teaching curriculum. Based on the teachers' estimation and the researcher's observation, the test takers should have A2 English listening ability according to the Common European Framework of Reference.

## 2.4 Procedure

The procedure of this study is illustrated in ***Graph 1***. Here I will focus on explaining several steps that are of most importance to this study.




*Graph 1: The procedure of this study*

### 2.4.1 Step 3: Speech sample recording

In step 3 of ***Graph 1*** the researcher recorded speech samples from the 12 speakers. The researcher prepared 12 scripts for speech samples, each of which consisted of approximately 60 words (+/- 3 words). The script was about

international affairs and the researcher ensured that there were no obscure words or phrases that might cause difficulty in comprehension for the listeners. Each speaker was asked to record a different script in 20 seconds (+/- 2 seconds). The reason for setting the 20-second time limit is based on previous research that reported a 20-second speech sample is of appropriate length for investigating issues such as accent strength and identification (Adank et al., 2009). It should be mentioned that 60 words in 20 seconds approximately equals 180 words per minute (WPM), which is faster than normal speech rate. The rationale for setting a faster rate is because during piloting the researcher found that potential speakers tended to polish their accents when they could read the scripts more slowly, which consequently reduced the accent strength and accent identifiability of their speeches. Therefore, after a few trials, the speech rate was set at 180 WPM. All 12 speakers were given sufficient time to familiarise themselves with the sample scripts and the final 20-second samples were quality-controlled to be ensure fluency and accuracy.

### 2.4.2 Step 5: Selecting four speakers for the Test

The results from the Task were analysed as follows:

***1.*** *The answers for accent strength were coded with numbers from 1 to 5 with 1 indicating a weak accent rating (Band 1) and 5 indicating a strong accent rating (Band 5).*

***2.*** *The answers for accent identification were coded as when a listener successfully identified the accent of the speaker, the speaker received a rating of 1 from that listener and when a listener failed, the speaker received a 0 rating.*

The mean scores of each speaker's accent strength and identifiability are reported in ***Chart 1*** and ***Chart 2***. The columns that are marked in purple refer to speakers that were selected for the Test recording.

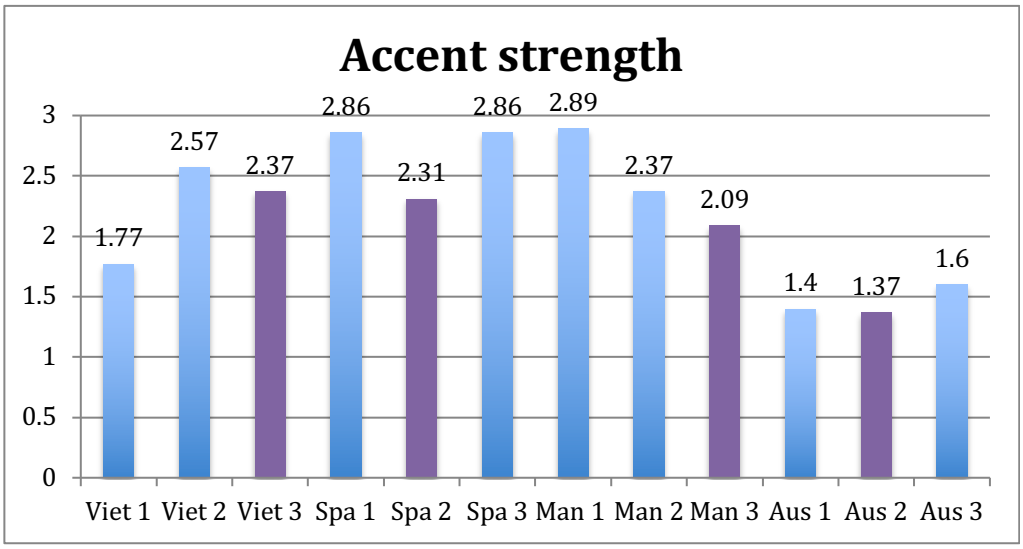



*Chart 1: Mean scores of speakers' accent strength*

*(Viet=Vietnamese, Spa=Spanish, Aus=Australian, Man=Mandarin)*

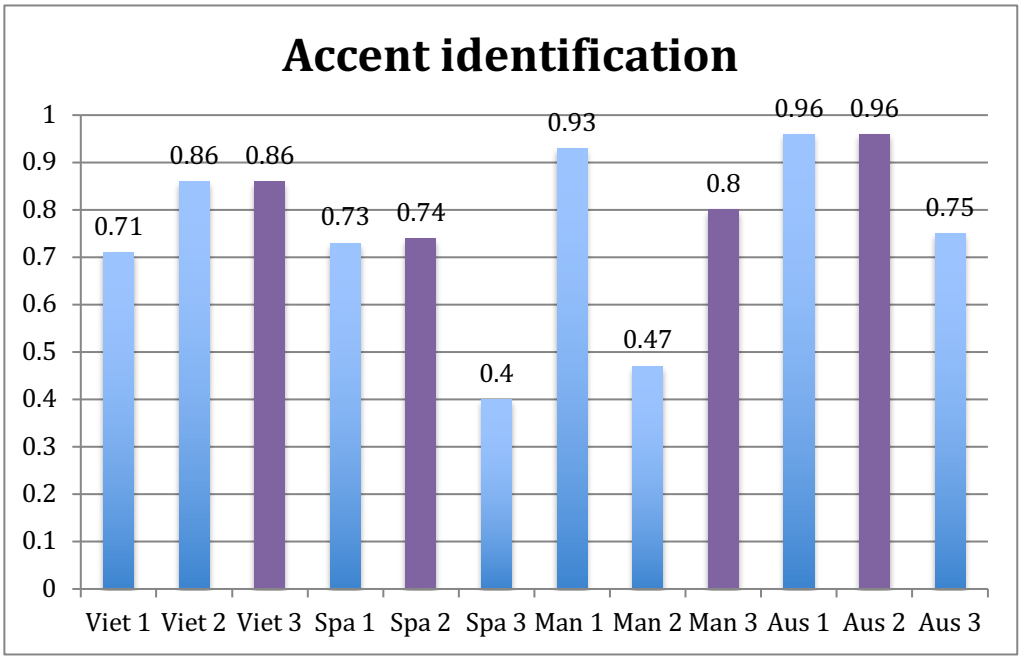



*Chart 2: Mean scores of speakers' accent identifiability*

The selection criteria for the four final speakers are as follows:

***1.*** *The accent strength of the four speakers should be within a similar range, ideally, 2.0-2.5, which, according to the Scale, was light-to-mild and suitable for young test takers based on the findings from the pilot study.*

***2.*** *The accent identifiability of the four speakers should be >0.7, which indicated that generally 70% of the listeners could successfully identify the speakers' first language.*

It should be noted the accent strength for the final Australian accent speaker, Aus 2, was only 1.37, which is much lower than the other three final

speakers. However the researcher still considered Aus 2's accent strength comparable to the other three. The reason is that all 35 listeners selected in this study were residing in Australia and had much more exposure to the Australian English accent. Due to high level of familiarity, they naturally found the Australian English accent more comprehensible and of a less accent strength in general as proven by the low accent strength rating for all three Australian speakers. Therefore, even though Aus 2's accent strength is approximately one band lower than other final speakers, the researcher still considered it to be at the same level as the other three's accent strength.

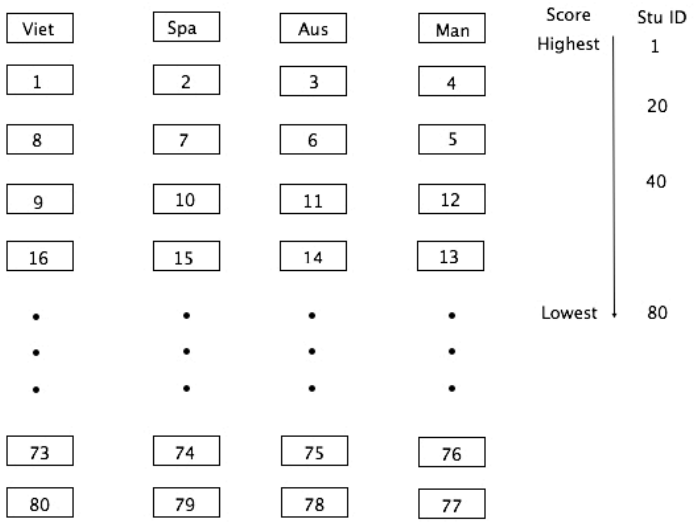
### 2.4.2 Step 6: Test recording

Test recording was conducted in a soundproof room with top-quality recording systems. The selected speakers were first given time to familiarise themselves with the Test scripts and then asked to practise reading the script in 5 minutes (+/- 10 seconds) so as to ensure all four accent recordings would be of a similar speech rate. During the recording, if the speakers misread certain segments, they were asked to reread the entire sentence. The reason for such a request is that the researcher had noticed during pilot recording that if speakers only rerecorded certain segments but not the complete sentence, the researcher needed to insert and edit these corrected segments into the misread sentences, which would disrupt the stress and intonation pattern of the whole sentence. Therefore, to ensure that their speeches would sound as natural as possible, whenever an error segment occurred, the speaker would be asked to reread the whole sentence where the segment originated instead of just the error segment.

The four finished recordings were edited through “Audacity”, an application that allowed the researcher to remove hesitations, repetitions, background noises and unexpected sounds in the recordings. Four volunteer test takers and four corresponding English teachers were asked to take the Tests (one student and one teacher per version) to ensure that the Test and the recordings worked well.

### 2.4.3 Step 7: Grouping test takers

There were 253 15-year-old students in that middle school at the same year level. 80 test taker were selected based on the scores from a 25-item preliminary English listening test that was administered to all 253 students. Out of all 253 students, the 80 students with the highest scores were chosen for this study. The researcher divided the 80 students into four groups with as minimal between-group difference as possible based on their test scores in the preliminary test. In order to achieve such a purpose, the researcher assigned each student with an ID number with the top-scoring student given 1 and the 80$^{th}$ scoring student given 80. The students were then divided into four groups according to the zigzag grouping method shown in ***Graph 2***.



*Graph 2: The grouping method*

In order to examine the between-group difference, a one-way ANOVA was conducted with the group as the factor and the scores from the preliminary listening test as the dependent variable. Descriptive statistics are reported in ***Table 3*** and show that the test score means of the four groups were highly similar. ANOVA analysis also revealed no significant between-group difference: $F_{(3,76)} = 0.006$, $p = 0.999$. Results indicated that the four groups were essentially identical in terms of English listening competence.

| Group | N | Mean | SD |
|---|---|---|---|
| **Vietnamese** | 20 | 17.90 | 2.65 |
| **Spanish** | 20 | 18.00 | 2.71 |
| **Mandarin** | 20 | 17.90 | 2.91 |
| **Australian** | 20 | 17.95 | 2.76 |

*Table 3: Descriptive statistics for the preliminary listening best*

### 2.4.4 Step 8: Administering the Test and the Questionnaire

After piloting all the research instruments with both coordinating teachers and volunteer student participants (who were not part of the main group of 80 test takers), the Test and Questionnaire were administered to the 80 test takers after school hours in the participating middle school in China.

The four groups of test takers were placed in four different classrooms, each of which was supervised by two corresponding English teachers. The Test recording lasted 5 minutes and since it was played twice to suit test takers' listening competence, the whole test finished in around 10 minutes. The Questionnaire then followed, which lasted around 13 minutes. It should be mentioned that before the this study commenced, ethics clearance was gained from all participating student test takers, students' parents and the principal of that middle school. Both student test takers and their parents were assured by their English teachers that test takers' performance in the Test or their answers in the Questionnaire would not affect the normal assessment in their English class. Especially with the Questionnaire, test takers were encouraged by the teachers to respond as honestly as possible instead of giving answers which they thought would please the teachers.

### 2.4.5 Step 9: Test scoring

To ensure reliable scores, the Test papers were marked consecutively by two different raters. The results from the Test and Questionnaire were coded and entered into SPSS spreadsheets by the researcher and double-checked by another volunteer to guarantee correctness. In terms of the marking rubric for the Test, both Construct 1 and 2 consisted of objectively scored items, which allowed for only one correct answer for each question. Construct 3 involved filling missing words based on contextual clues so any answers that were contextually acceptable were given full marks. Grammatical mistakes were penalised with no marks given.

# Chapter 3. Data analysis and results

## 3.1 RQ 1: ANOVA on the Test scores

A reliability analysis was run on the 30 items of the Test and the result was $\alpha$=0.689, which is reasonable for a test of little stakes. A one-way ANOVA was administered on the Test scores with the accent group as the factor and the test scores as the dependent variable. ***Table 4*** and ***Chart 3*** illustrate the descriptive statistics, which show that the Mandarin group scored the highest, followed by the Spanish, Australian and Vietnamese groups. There is also conspicuous score difference between the Mandarin and the other three groups, with Mandarin leading a near 2.5-point difference from the second highest. By contrast, the differences among the three non-Mandarin groups are much smaller, with only a one-point difference between each group. ANOVA results offer solid evidence that there is significant difference between groups: $F_{(3,76)} = 6.844$, $p < 0.001$, $\eta^2 = 0.21$. Though the overall effect size is small to medium, it should be understood that the indication here is that 21% of the test score difference is caused by the accent of the test, which shows that accent can influence test takers' performance strongly.

| Group | N | Mean | SD |
|---|---|---|---|
| **Vietnamese** | 20 | 17.70 | 3.34 |
| **Spanish** | 20 | 19.65 | 3.54 |
| **Mandarin** | 20 | 22.05 | 3.49 |
| **Australian** | 20 | 18.50 | 2.46 |

*Table 4: Descriptive statistics for the Test*

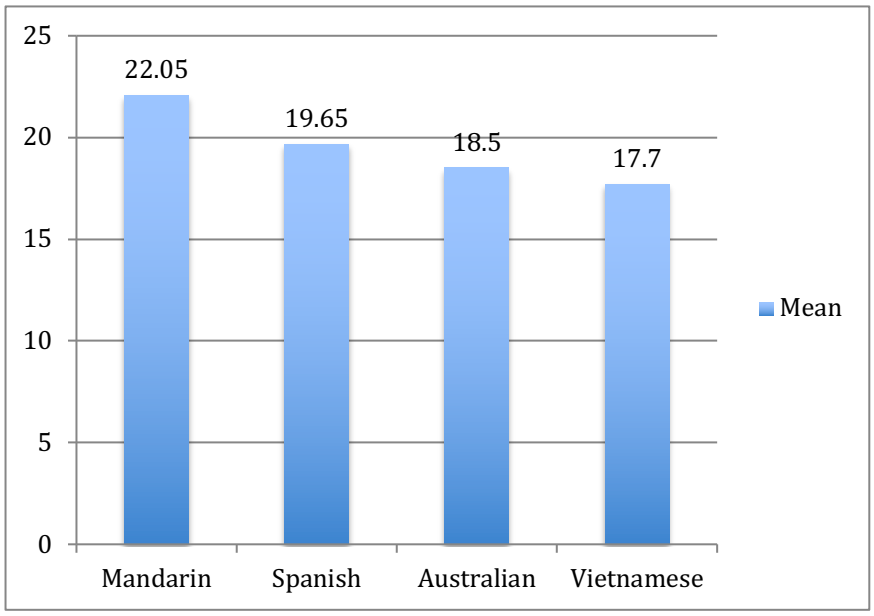



*Chart 3: Group mean scores in the Test*

A Scheffé Post Hoc test was conducted and between-group significance levels and effect sizes are reported in ***Table 5***. The Mandarin group is significantly different from the Vietnamese and Australian groups. Though the difference between the Mandarin and the Spanish is not significant, it is observed that the significance level for the Mandarin and Spanish comparison (p=0.149) is still much smaller than the ones among the three non-Mandarin comparisons (p=0.894, 0.312, 0.739). Large effect sizes were observed in the Mandarin-Vietnamese (d=1.273) and the Mandarin–Spanish (d=1.176) pair. There is also a medium to large effect size between Mandarin and Spanish (d=0.683).

| Accent groups | | Significance level p | Effect size d |
|---|---|---|---|
| **Mandarin** | **Vietnamese** | 0.001 | 1.273 |
| | **Australian** | 0.011 | 1.176 |
| | **Spanish** | 0.149 | 0.683 |
| **Vietnamese** | **Australian** | 0.894 | 0.273 |
| | **Spanish** | 0.312 | 0.567 |
| **Australian** | **Spanish** | 0.739 | 0.377 |

*Table 5: Post Hoc test on the Test results*

### 3.2 RQ 2: ANOVA on the Questionnaire sub-scales

The researcher adopted a quantitative approach to analyse the Questionnaire data where answers to the Likert-scale items were coded with numbers from 1 to 7. The coding principle is that a higher number should always be given to greater familiarity, stronger comprehension and a more positive attitude. Therefore, an answer of "strongly agree" was coded with 7 and "strongly object" with 1. Four negatively phrased items (item 5, 8, 11, 15) were reverse coded. The item results for each sub-scale were totaled and treated as interval data. The reliability of the whole Questionnaire including all 15 items was $\alpha$=0.794 while separately Familiarity was $\alpha$=0.636, Comprehension $\alpha$=0.742 and Attitude $\alpha$=0.676. The high overall and sub-scale reliabilities imply that there is a strong relationship between all three sub-scales.

To further investigate the between-group difference of test takers' subjective perception, one-way ANOVA was conducted separately on Familiarity, Comprehension and Attitude.

### 3.2.1 Familiarity

Descriptive statistics for Familiarity are presented in ***Table 6*** and ***Chart 4.*** There is no significant between-group difference as determined by ANOVA: $F_{(3,76)} = 1.816$, $p = 0.151$. Descriptive statistics show that test takers found the Spanish accent most familiar (mean=22.75), followed by Mandarin and Australian (mean=20.45) as the second and Vietnamese as the least (mean=18.90).

| Group | N | Mean | SD |
|---|---|---|---|
| **Vietnamese** | 20 | 18.90 | 4.15 |
| **Spanish** | 20 | 22.75 | 3.81 |
| **Mandarin** | 20 | 20.45 | 6.05 |
| **Australian** | 20 | 20.45 | 6.53 |

*Table 6: Descriptive statistics for Familiarity*

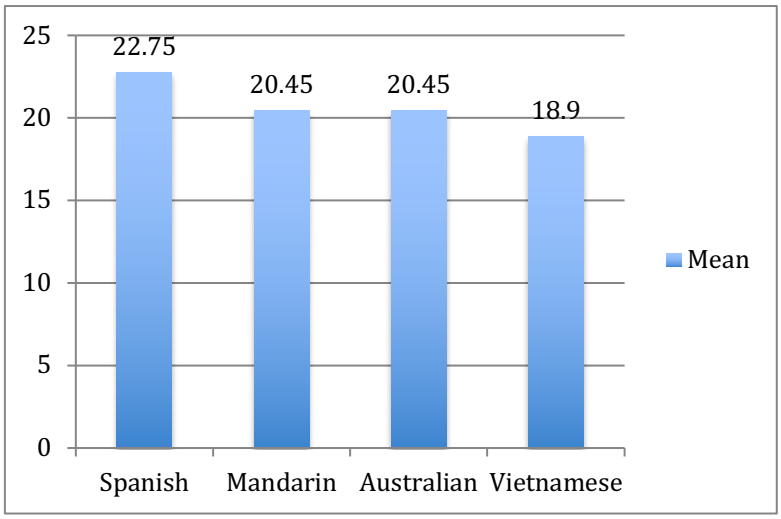



*Chart 4: Group mean scores in Familiarity*

### 3.2.2 Comprehension

***Table 7*** and ***Chart 5*** illustrate the descriptive statistics for Comprehension, which shows that test takers found the Mandarin accent most comprehensible (mean=22.80), followed by Spanish (mean=20.32), Australian (mean=19.49) and

Vietnamese (mean=18.00). Interestingly just like the Test results, the mean score of the Mandarin group in Familiarity is also much higher than the second group by a near 2.5-point difference. In comparison, the difference among the three non-Mandarin groups is much smaller, which is generally a one-point difference.

| Group | N | Mean | SD |
|---|---|---|---|
| **Vietnamese** | 18 | 18.00 | 3.63 |
| **Spanish** | 19 | 20.32 | 4.53 |
| **Mandarin** | 20 | 22.80 | 5.05 |
| **Australian** | 20 | 19.49 | 5.44 |

*Table 7: Descriptive statistics for Comprehension*

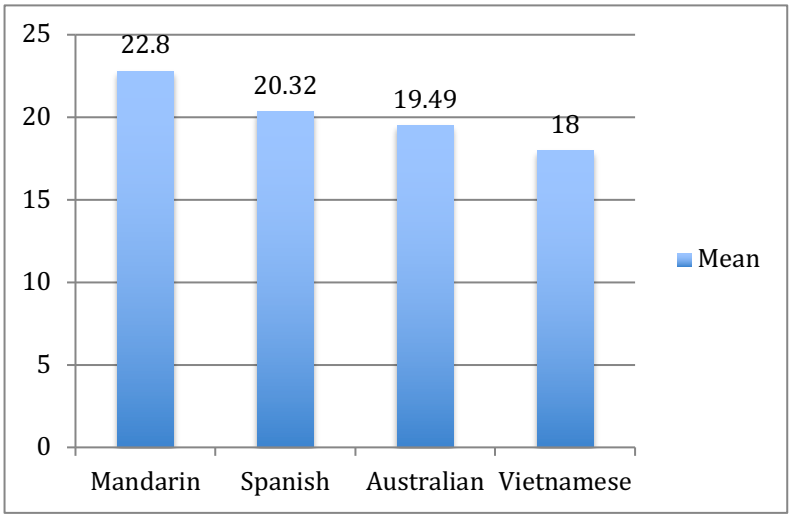



*Chart 5: Group mean scores in Comprehension*

ANOVA reveals that the between group difference is significant, $F_{(3,73)}$ = 6.260, p =0.001, $\eta^2$=0.20. A Scheffé Post Hoc test was conducted and the results are reported in ***Table 8***. There is significant difference in the Mandarin-Vietnamese pair (p=0.027, d=1.091) and the Mandarin-Australian pair (p=0.002, d=0.631) and the effect sizes for the two pairs are large. No significant difference is found in the Mandarin-Spanish pair or any of the non-Mandarin pairs.

| Accent groups | | Significance level p | Effect size d |
|---|---|---|---|
| **Mandarin** | **Vietnamese** | 0.027 | 1.091 |
| | **Australian** | 0.002 | 0.631 |
| | **Spanish** | 0.450 | 0.517 |
| **Vietnamese** | **Australian** | 0.883 | 0.322 |
| | **Spanish** | 0.534 | 0.565 |
| **Australian** | **Spanish** | 0.148 | 0.166 |

*Table 8: Post Hoc test on Comprehension*

### 3.2.3 Attitude

Descriptive statistics for Attitude are presented in ***Table 9*** and ***Chart 6*****.** Test takers from the Spanish group responded most positively to their accent (mean=26.35), followed by Mandarin (mean=25.85), Australian (mean=24.60) and Vietnamese (mean=24.50). There is a less than two-point difference between the most favoured and the least favoured accents so in general there is little difference in terms of test takers' attitude to different accents. ANOVA analysis further confirms such an observation because the difference is found to be non-significant: $F_{(3,76)} = 0.668$, $p = 0.574$.

| Group | N | Mean | SD |
|---|---|---|---|
| **Vietnamese** | 20 | 24.50 | 4.77 |
| **Spanish** | 20 | 26.35 | 4.94 |
| **Mandarin** | 20 | 25.85 | 6.07 |
| **Australian** | 20 | 24.60 | 4.12 |

*Table 9: Descriptive statistics for Attitude*

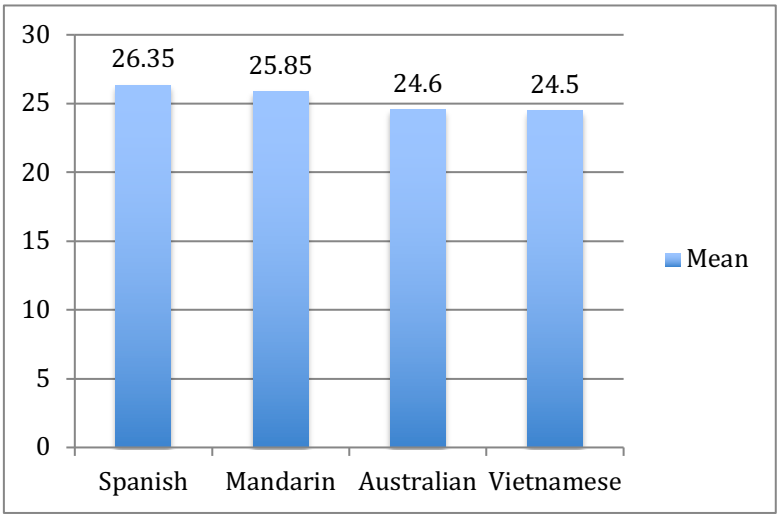



*Chart 6: Group mean scores in Attitude*

### 3.3 RQ 3: Correlation of Questionnaire sub-scales with Test scores

Pearson correlation was run on the three Questionnaire sub-scales with the Test scores, the findings of which are presented in ***Table 10***. All three sub-scales correlate positively and significantly with Test scores while the strength of correlation varies, with Comprehension being the strongest (r=0.527) followed by Attitude (r=0.389) and Familiarity (r=0.228). The strong correlation between Comprehension and test scores indicates that test takers' subjective perception of the comprehensibility of the test can largely reflect their performance in the Test, with better self-perceived Comprehension related to higher Test scores. By comparison, test takers' attitude and familiarity towards accents are not strong predictors of their test performance.

| Correlation | | N | Pearson's r | Significance level p |
|---|---|---|---|---|
| **Test scores** | **Familiarity** | 80 | 0.228 | 0.042 |
| | **Comprehension** | 80 | 0.527 | <0.001 |
| | **Attitude** | 80 | 0.389 | <0.001 |

*Table 10: Correlation of Questionnaire sub-scales with Test scores*

### 3.4 RQ 4: Correlation among the three Questionnaire sub-scales

Since the sub-scales were considered interval data, Person's r was used to examine the correlation among them and the results are reported in ***Table 11***. The findings indicate that both Familiarity and Comprehension correlate positively and significantly ($p<0.001$) with Attitude. The strength of correlation for these two pairs is considered to be very strong (Familiarity-Attitude, $r=0.497$; Comprehension-Attitude, $r=0.512$), given that a listener's attitude towards an accent can be affected by many potential factors. Familiarity and Comprehension almost correlate significantly ($p=0.077$) and the correlation is medium ($r=0.203$).

| Correlation | | N | Pearson's r | Significance level p |
|---|---|---|---|---|
| **Familiarity** | **Comprehension** | 80 | 0.203 | 0.077 |
| | **Attitude** | 80 | 0.497 | <0.001 |
| **Comprehension** | **Attitude** | 80 | 0.512 | <0.001 |

*Table 11: Correlation among three Questionnaire sub-scales*

# Chapter 4. Discussion and Conclusion

## 4.1 Answer to RQ 1: The difference between test takers' performance in the Test

As reported in **Section 3.1**, the Mandarin accent group performed significantly better than the Vietnamese and the Australian groups with substantial effect sizes in both pairs. The Mandarin group scored almost significantly better than the Spanish group with a considerable effect size as well. This lends strong support for the shared-L1 advantage phenomenon noticed in Harding (2011) and Major et al. (2002). The large effect sizes reported in this study further highlight the extent to which a shared-L1 background can advantage test takers' performance at a global listening test level. Though the difference in the Mandarin-Spanish pair is not significant and not as pronounced as the other two pairs, the Mandarin-Spanish pair certainly demonstrates a similar pattern to the other two pairs. The reason for such an argument is that though the significance level for the Mandarin-Spanish pair is comparatively bigger than the other two pairs, it is still much smaller than the levels among the three non-Mandarin pairs. The effect size for the Mandarin-Spanish pair (d=0.683) should also be interpreted along with the significance level and such a medium to large effect size does prove that the score difference in the Mandarin-Spanish pair is of considerable magnitude. Therefore there is a tendency for the difference in the Mandarin-Spanish pair to be just as noteworthy as the other two and the possible reason for the lack of significance might just be a case of small sample size as there were only 20 participants for each group.

Another interesting finding from ANOVA is that there is no significant difference between the three non-Mandarin groups, which indicates that test takers found the non-Mandarin accented Tests of similar difficulty, whether it is a native accent (Australian), a non-tonal non-native accent (Spanish) or a tonal non-native accent (Vietnamese). The implication here is thought-provoking in that if such a finding is generalisable, it implies test takers do not perform significantly differently with various English accents, native or non-native, phonologically similar to their L1 or phonologically different from their L1, as long as this accent is not their L1 accent. This can reassure test designers and testing organisations that the inclusion of non-native English accents would not advantage or disadvantage any particular test taker groups in terms of test scores provided they could circumvent the shared-L1 effect issue.

## 4.2 Answer to RQ 2: The difference between test takers' perception of accents in the Questionnaire

### 4.2.1 Comprehension

The results for ANOVA on the Comprehension sub-scale in ***Table 7*** and ***Table 8*** largely mirror the findings from ANOVA on the Test scores in ***Table 4*** and ***Table 5***. Not only did the Mandarin group perform significantly better than the other three non-Mandarin groups in the Test, the Mandarin group also judged the Mandarin accent more comprehensible than the other three groups at a significant level. This result is unsurprising in that it is expected that if test takers found a particular accent more comprehensible, they would perform better in that accented test and vise versa. Here the Comprehension scale in essence investigated the same issue as the listening test and successfully offered

an emic perspective to triangulate the findings from the psychometric measurement. Test takers' judgment of the comprehensibility of the three non-Mandarin accents also corresponds with their performance in the test with Spanish ranked the second, followed by Australian and Vietnamese. It should be noted that in this study the native accent, Australian accent, neither facilitated test takers' performance in the test, nor was it considered more comprehensible or intelligible from test takers' subjective perception in the Comprehension scale.

#### 4.2.2 Familiarity and Attitude

As reported in **Section 3.2.1** and **3.2.3**, ANOVA failed to find any significant between-group difference in both Familiarity and Attitude. Such findings indicate that test takers did not find any particular accent more familiar or more favourable. The result for Attitude particularly challenges the established argument that listeners prefer the native accents of a certain language to non-native accents (Fraser & Kelly, 2012; Hendriks, Meurs & Groot, 2015; Hendriks, Meurs & Meji, 2015). Here the researcher proposes two explanations for such a finding.

Firstly, the researcher did not inform test takers of the nationalities of the speakers so test takers would not get influenced by any accent-irrelevant factors. Secondly, the test takers selected for this study were young learners who had little exposure to different English accents and they were not expected to have formed any significant preconceived judgments over the three non-Mandarin accents in this study. Therefore these young test takers would be less concerned with issues such as the social status or prestige centering around particular accents as was the case with the British accent in Nejjari, Gerritsen, Van der

Haagen and Korzilius (2012). Due to the lack of experience with English accents, test takers in this study were also assumed to judge the familiarity and favouribility of the four accents based largely on their intuitive understanding, which circumvented issues such as distorted accent perception (Hu & Lindemann, 2009) and perception bias resulting from accent identification (Winke & Gass, 2013).

Even with their shared-L1 accent (Mandarin accent), the test takers did not seem to have successfully identified it as they considered the Spanish accent more familiar than the Mandarin accent. It is indeed highly unlikely for them to have had more exposure to the Spanish accent than the Mandarin accent so a possible explanation could be that low-proficiency test takers lacked the phonological sensitivity to tell the Mandarin accent apart from other non-Mandarin accents. They might have rated Spanish accent more familiar simply because they liked this accent, as Spanish accent is rated the most favourable on the Attitude scale and there is a strong correlation between Familiarity and Attitude in general. Therefore, when issues such as accent identification (McKenzie, 2015) and categorization (Atagi & Bent, 2014) are removed from the accent judgment process as well, the take-home message from the findings in Familiarity and Attitude is clear: inexperienced listeners perceive various English accents as equally familiar and favourable when they were not influenced by any extraneous factors.

This finding in the Attitude scale has also confirmed what Chien (2014) noticed in that when not informed of the nationalities of speakers, Chien's listeners did not show any preference for native English accents such as Australian accent or British accent. Though a positive attitude to American

English accent was observed in Chien (2014), this is largely due to the fact that American English accent has been used as the standard accent for EFL teaching in Taiwan and English learners had extensive exposure to it.

A similar argument can also be made for the findings in Butler (2007) when she reported that her young Korean English learners demonstrated a more positive attitude towards the American accent than the Korean accent even though both accents were equally comprehensible. What Butler (2007) did not mention in her study is that the standard American English accent has also been adopted as the major accent input in Korean English teaching and testing settings, which was reported in Yook and Lindemann (2013).

Therefore, the partiality towards certain dominant native English accents noticed in previous research (Butler, 2007; Chien, 2014) should not be taken as evidence that native English accents are somehow inherently more favourable than non-native accents for English learners. Such a preference might simply be caused by pedagogical practices that overly focus on certain accents to the detriment of creating biased accent perception in learners. As is statistically evidenced in **Section 3.4**, high familiarity with an accent is indeed strongly correlated with a more positive attitude.

## 4.3 Answer to RQ 3: The relationship between Test results and Questionnaire results

### 4.3.1 Correlation between Comprehension and Test scores

***Table 10*** illustrates the connection between the etic perspective and emic perspective of how test takers responded to the four accents. The strong correlation between test scores and Comprehension scores has been explained

in **Section 4.2.1** and here the researcher will focus on the correlations in the Test scores-Familiarity and Test scores-Attitude pairs.

### 4.3.2 Correlation between Familiarity and Test scores

Compared to the strong correlation in the Test scores-Comprehension pair, the correlations in the other two pairs are much weaker. The weak correlation between Test scores and Familiarity indicates that although a test taker may find a particular accent very familiar, familiarity does not translate strongly into better performance in the test.

This seems to contradict the findings in Ockey and French (2014) when they reported a general pattern where familiarity with an accent aided test performance. However, this pattern should still be considered as a tentative argument since even Ockey and French reported evidence that did not fully support this argument. Firstly at an individual speaker level in their study, there was one speaker with a light Australian accent and test takers who reported better familiarity with Australian accent actually performed worse than test takers who reported less familiarity in this speaker's case. Second, at an accent group level in their study, the relationship between familiarity and test scores was also inconsistent between the British accent speaker group and the Australian accent speaker group.

Ockey and French attributed such discrepancies to the fact that self-perceived familiarity with accents could be unreliable and this argument has been supported by the findings in this study as well. It is implausible to suggest the test takers in the researcher's study would be more familiar with Spanish

than with Mandarin English accent as reported in ***Chart 4***. Herewith the researcher's explanation is that my test takers were actually more familiar with Mandarin accent as a result of previous extensive exposure to Mandarin-accented English from their English teachers who are all L1-Mandarin speakers. Repeated exposure to their L1 accent improved test takers' comprehension of L1 accented recordings and this translated into better test performance, which is the shared-L1 effect. However, when asked to subjectively rate their familiarity with different accents, test takers may lack the phonological sensitivity to accurately identify an L1 accent, especially in the case of young test takers. This point has already been discussed in **Section 4.2.2.**

Another possible explanation is that accent familiarity itself is a slippery concept. Test takers might perceive an L1 accent of a particular novel speaker as unfamiliar because they have never heard this person speaking before. However, phonologically test takers might be familiar with the segmental and suprasegmental features in this speaker's accent and such "unconscious familiarity" could improve their comprehension of a listening test recorded in this speaker's accent.

### 4.3.3 Correlation between Attitude and Test scores

The strength of the correlation between Test scores and Attitude is weak, which implies that a test taker's attitude towards an accent can to a certain extent influences their performance in a test of that accent, though the relationship is not pronounced. Besides, when looking at the insignificant between-group difference from ANOVA on the Attitude scale, it shows that the young test takers in this study did not have any particular preference for any of the four accents at

a group level. When the two findings are interpreted together, it indicates that test takers' attitude towards accented listening input was of little variance and their attitude would only have a very limited impact on their test performance as well. This is actually what test makers would hope for because test takers' attitude to accents should not affect their performance. A sound listening test should only measure test takers' listening ability and attitude would be considered as a construct-irrelevant variance.

### 4.4 Answer to RQ 4: The relationship among the three sub-scales in the Questionnaire

Findings in ***Table 11*** demonstrate that there are strong and significant correlations in the Familiarity-Attitude pair and the Comprehension-Attitude pair. The relationship here should be understood as if a test taker found a particular accent more familiar or more comprehensible, they were likely to develop a more positive attitude towards it. The finding in the Comprehension-Attitude pair is in agreement with what Scales, Wennerstrom, Richard and Wu (2006) noticed when their learners considered the most comprehensible accent to be the most preferred accent and their correlation was near perfect. The finding in the Familiarity-Attitude pair furthers the argument in Scales, Wennerstrom, Richard and Wu (2006) in that it points out that language learners' familiarity with an accent can have a similarly strong predicting power on their attitude of that accent just as the Comprehension scale.

The interesting finding here is that Familiarity and Comprehension do not correlate as strongly and significantly as the other two pairs. This seems to contradict what most previous research has reported when they argued that

more familiarity with an accent could lead to stronger comprehension of listening materials of that accent (Adank et al., 2009; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009; Weber, Broersma & Aoyagi, 2011). However it should be noted that the above-mentioned studies come from a neuroscience perspective and comprehension was mostly defined as measurable objective constructs such as processing speed as in Adank et al. (2009). By comparison, Comprehension in this study was a questionnaire construct and measured test takers' subjective reflection on how much they thought they could understand a particular accent. Therefore the weak correlation in the Familiarity-Comprehension pair should be interpreted as a test taker might find a certain accent very familiar in the questionnaire but they did not report high confidence in comprehending it. On the other hand, a test taker might consider they could comprehend a particular accent very well but that accent might not necessarily be an accent that they found highly familiar.

Another explanation is the slippery definition of Familiarity as discussed in **Section 4.3.2**. Though the researcher stressed "the accent of the test" in the Questionnaire, test takers might still misinterpret "familiarity with an accent" as "familiarity with a particular speaker". A possible scenario is that a listener indeed found an accent more comprehensible because they were more familiar with this accent. However, because they had never heard this speaker speaking before, they rated this accent as unfamiliar. Therefore, the weak correlation between Familiarity and Comprehension is understandable if we consider how Familiarity and Comprehension were constructed in this study and how they might be interpreted by test takers.

## 4.5 Implications

Based on answers to the four research questions, the researcher will discuss the implications drawn from this study in three directions, the implications for language testing, language teaching and stakeholders.

### 4.5.1 Implications for language testing

Findings from this study have offered support for the shared-L1 advantage argument and if this issue can be addressed properly, there is possibility for the inclusion of non-native English accents into high-stakes international English listening assessments. It is clear from the answers to RQ1 and RQ2 that test takers were advantaged by their shared-L1 accent and they also identified their shared-L1 accent as the most comprehensible one. Therefore, both the etic and emic perspectives have shown that a shared-L1 accent can improve test takers' comprehension of the listening test materials. Possible solutions for tackling the shared-L1 issue have been suggested in Harding (2011) such as only including the most frequently used non-native accents in a given context, adopting highly intelligible non-native accents and balancing non-native accents with various listening tasks. However, those suggestions are still tentative and further research is needed to validate them.

On the other hand, the answers to RQ1 and RQ2 can also be interpreted as solid support for multidialectal listening assessments in that there is no statistical difference in either test takers' performance or their evaluation of Comprehension in the three non-L1 accents. The implication here is that if an international listening test is to include various English accents, the shared-L1 effect might be the only major issue test makers need to address in order to

preserve test fairness and avoid test bias. The three non-L1 accents selected for this study are highly representative in that for the L1-Mandarin test takers, Australian accent is a native English accent, Vietnamese accent is a tonal non-native accent and Spanish accent is a non-tonal non-native accent. Since no significant difference can be observed among the three non-L1 accents, it is reasonable to expect that perhaps test takers would demonstrate similar comprehensibility in a listening test as long as the accent is not their L1 accent. Such a possibility can greatly simplify the selection of accents in listening tests and though more studies are clearly needed before such an argument can be ascertained, this certainly heralds a promising research direction.

From the emic perspective, the findings from RQ3 can also be interpreted as supporting the practice for including non-native accents in listening tests in that they indicate that test takers' performance in multidialectal assessment can be largely predicted by their comprehension of the listening recording, which is exactly what a listening test construct should aim for. The answers to RQ3 show that test taker's Familiarity and Attitude have very limited impact on their performance in the test, which is precisely what test designers would desire since accent Familiarity and Attitude would be considered construct-irrelevant variances. Therefore, regardless of test taker's familiarity with or attitude to the accented recordings, a multidialectal listening test can still measure test taker's genuine listening competence. However, a cautionary note should be offered that the suggestions here are based on the fact that only light-to-medium accents were adopted. The impact of accent strength on test takers' performance was not examined in this study.

### 4.5.2 Implications for language teaching

The findings from RQ 4 have implications for language teaching in the ELF context. As Canagarajah (2006) argues, English users today need to possess the ability to shuttle between and negotiate with different English varieties. Therefore, English classes and English teachers need to prepare learners to face multidialectal English communication with confidence, competence and a positive attitude. It would be difficult for learners to communicate effectively with accented English speakers if they have a preconceived negative attitude towards dialectal or non-native accents. The optimistic news here is that the findings from RQ4 show that young learners' attitude towards English accents can be largely shaped by their familiarity with that accent and comprehension of that accent. English teachers and textbook designers can take advantage of such findings and expose young learners with more less-dominant native and non-native English accents. It is indeed possible that learners may dislike an accent if they are just not familiar with it, not used to listening to it or simply find it hard to understand. The fact that existing English textbooks and test recordings in EFL contexts overly focus on one or two more prestigious native accents can only exacerbate this situation because learners may think only these few accents they are familiar with and can easily comprehend are the "good" accents or "standard" accents.

Therefore test designers and English teachers should take up the responsibility to inform learners that English accent is not a homogenous concept centering on one or two accents. They need to cultivate an awareness in learners that allows them to appreciate various L1 and L2 English accents. The results from RQ4 therefore point out two variables that educators can work on to

develop learners' readiness for the multidialectal reality of English. Educators can try to introduce more accent varieties into textbook recordings so that learners can gain better familiarity with different accents. Specific tasks and activities can also be designed to improve learners' ability of comprehending diverse English accents in classroom settings. Such practices can to a large extent help learners shape a more positive attitude when they now live in a world where multidialectal English communication skills are of vital importance.

### 4.5.3 Implications for stakeholders

When a new language test is put into the market, it is understandable that test makers and testing companies want the test to be popular and commercially successful. Therefore, aside from ensuring the test is well designed, test makers also need to make sure that the test will be welcomed by stakeholders, part of whom are test takers. If test takers found the newly introduced non-native accents unpleasant, they might perceive the test negatively in general and refuse to take it anymore. However, the good news is that the findings from RQ2 reveal that test takers found the four accents in the test of similar familiarity and displayed a neutral, non-discriminative attitude to them. In fact, the test takers in this study even favoured the Spanish accent (a non-native accent) more than the Australian accent (a native accent) as indicated in ***Chart 6***.

To push this argument even further, if the shared-L1 effect on Comprehension is properly handled, the answers to RQ2 basically indicate that young test takers' subjective perception of the four accents was impartial and indifferent, whether by the measurement of Familiarity, Comprehension or Attitude. Such a finding would be of great value to test makers and testing

companies because it implies that the move to multidialectal listening assessment would carry little commercial risk since test takers would just welcome the new listening test as much as they would do with the old ones. Therefore, the introduction of non-native accents is not only conceptually legitimate based on the ELF argument, but also will be a practice positively perceived by test takers, hence maintaining a listening test's popularity and practicality.

### 4.6 Conclusion and limitations

In conclusion, this study is situated in the ELF debate and examined from both the etic and the emic perspectives how young test takers performed in a listening test when non-native accents were introduced. The main finding is that there is a strong shared-L1 effect with Mandarin-L1 test takers who listened to the Mandarin accented test recording both in the Test and in the Questionnaire Comprehension sub-scale. Such an advantage in a testing setting would be considered test bias and test makers would need to tackle this issue if non-native accents are to be introduced. On the other hand, there is no significant difference in test takers' performance in the three non-Mandarin accents or their perception of the four accents in terms of Familiarity and Attitude, which supports the testing and teaching of multidialectal listening competence. The findings from this study have mixed implications for language testing, language teaching and relevant stakeholders but in general they support the practice of multidialectal listening assessment.

There are several limitations in this study. Firstly it only recruited test takers of one particular L1 background. Time and resource permitting, it would

be of greater value to recruit test takers of other L1s, such as Spanish or Vietnamese, so that the researcher could investigate if the shared-L1 advantage observed in the Mandarin-L1 group would also hold true with other L1 groups. Secondly, this study used one-way ANOVA to investigate global accent effect but did not analyse how test takers performed on individual items. Aside from examining the effect of accents at a test level, it would also be useful to investigate how the shared-L1 advantage manifests itself at an item level as investigated by Harding (2011) so that test makers could better identify the specific types of items or listening specifications that are most likely to induce the shared-L1 effect. When both the test level and item level are addressed, researchers can offer a more comprehensive picture for test makers and testing companies. Future research could focus on addressing these issues and broaden the research base in multidialectal listening assessment.

## Reference


Abeywickrama, P. (2013). Why not non-native varieties of English as listening comprehension test input?. *RELC Journal, 44*(1), 59-74.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 520-529.

Atagi, E., & Bent, T. (2013). Auditory free classification of native and nonnative speech by nonnative listeners. *Journal of Phonetics*, *41*, 1-23.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.

Brown, J. D. (2014) The future of world Englishes in language testing. *Language Assessment Quarterly, 11*(1), 5-26.

Butler, Y. G. (2007). How are nonnative-English-speaking teachers perceived by young learners?. *Tesol Quarterly, 41*(4), 731-755.

Cambridge English. (2015). Test format of IELTS (International English Language Testing System). Retrieved from http://www.cambridgeenglish.org/exams/ielts/whats-in-the-test/#

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly: An International Journal, 3*(3), 229-242.

Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal*, *91*, 923-939.

Chien, S. (2014). Varieties of English: Taiwanese attitudes and perceptions. *Newcastle and Northumbria working papers in linguistics, 20*, 1-16.

Clyne, M., & Sharifian, F. (2008). English as an international language: challenges and possibilities. *Australian Review of Applied Linguistics, 31*(3), 1-16.

Dai, W., March, D., & Victor, J. (2014) *The effect on accent on listening exercises in listening tests.* Unpublished manuscript, The University of Melbourne, Melbourne, Australia.

ETS. (2015). TOEFL iBT® Test Content. Retrieved from https://www.ets.org/toefl/ibt/about/content/

Fraser, C., & Kelly, B. (2012). Listening between the lines: social assumptions around foreign accents. *Australian Review of Applied Linguistics, 35*(1), 74-93.

Harding, L. (2011). *Accent and listening assessment: a validation study of the use of speakers with L2 accents in academic English listening assessment*. Frankfurt: Peter Lang.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29*(2), 163-180.

Hendriks, B., Meurs, F., & Groot, E. (2015). The effects of degrees of Dutch accentedness in ELF and in French, German and Spanish. *International Journal of Applied Linguistics*. doi: 10.1111/ijal.12101.

Hendriks, B., van Meurs, F., & van der Meij, E. (2015). Does a foreign accent sell? The effect of foreign accents in radio commercials for congruent and non-congruent products. *Multilingua, 34*(1), 119-130.

Hiraga, Y. (2005). British attitudes towards six varieties of English in the USA and Britain. *World Englishes, 24*(3), 289-308.

Hu, G., & Lindemann, S. (2009). Stereotypes of Cantonese English, apparent native/non-native status, and their effect on non-native English speakers' perception. *Journal of Multilingual and Multicultural Development*, *30*(3), 253-269.

Jenkins, J. (2006). The spread of EIL: A testing time for testers. *ELT journal, 60*(1), 42-50.

Jenkins, J. & Leung, C. (2013). English as a Lingua Franca. In A. J., Kunnan (Eds.), *The Companion to Language Assessment* (pp. 1605-1616). West Sussex, UK: John Wiley & Sons, Inc.

Kim, Y. S. (2007). *Korean adults' attitudes towards varieties of English*. Unpublished manuscript, University of Edinburgh, Edinburgh, United Kingdom.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL quarterly, 36*(2), 173-190.

Major, R. C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. (2005). Testing the effects of regional, ethnic, and international dialects of English on listening comprehension. *Language learning, 55*(1), 37-69.

McKenzie, R. M. (2008). Social factors and non-native attitudes towards varieties of spoken English: a Japanese case study. *International Journal of Applied Linguistics 18*(1), 63-88.

McKenzie, R. M. (2015). The sociolinguistics of variety identification and categorisation: free classification of varieties of spoken English amongst non-linguist listeners. *Language Awareness*, *24*(2), 150-168.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 111-131.

Nejjari, W., Gerritsen, M., Van der Haagen, M., & Korzilius, H. (2012). Responses to Dutch-accented English. *World Englishes, 31*(2), 248-267.

Ockey, G. J., & French, R. (2014). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, 1-24.

Papageorgiou, S., Xi, X., Morgan, R., & So, Y. (2015). Developing and validating band levels and descriptors for reporting overall examinee performance. *Language Assessment Quarterly*, *12*(2), 153-177.

Scales, J., Wennerstrom, A., Richard, D., & Wu, S. H. (2006). Language learners' perceptions of accent. *Tesol Quarterly, 40*(4), 715-738.

Seargeant, P. (2012). Disciplinarity and the study of world Englishes. *World Englishes*, *31*(1), 113-129.

Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Sharifian, F. (2013). Globalisation and developing metacultural competence in learning English as an International Language. *Multilingual Education*, *3*(1), 1-11.

Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America, 125*(5), 3306-3316.

Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The “other-accent” effect in voice recognition. *Journal of Cognitive Psychology, 24*(6), 647-653.

Stibbard, R. M., & Lee, J. I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433-442.

Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, *60*(1), 51-60.

Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes, 10*(2), 89-101.

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics, 39*(4), 479-491.

Weinberger, S. (2012). *Speech Accent Archive*. VA: George Mason University. Accessible at: http://accent.gmu.edu

Winke, P., & Gass, S. (2013). The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, *47*(4), 762-789.

Yook, C., & Lindemann, S. (2013). The role of speaker identification in Korean university students' attitudes towards five varieties of English. *Journal of Multilingual and Multicultural Development, 34*(3), 279-296.

Zhang, Q. (2009). Hong Kong people's attitudes towards varieties of English. *Newcastle Working Papers in Linguistics 15*, 151-173.

## Appendix A: The Strength of Accent Scale

**Band 1**
The speaker's accent is **very similar to** what I am used to.
I can concentrate on listening **without any problem**.
I can **easily** understand the recording.

**Band 2**
The speaker's accent is **slightly different** from what I am used to.
I can concentrate on listening **without too much trouble**.
I can understand the recording **to a large extent**.

**Band 3**
The speaker's accent is **different** from what I am used to.
I find it **slightly challenging** to concentrate on listening.
I can **roughly** understand the recording.

**Band4**
The speaker's accent is **very different** from what I am used to.
I need to concentrate on listening **more than usual**.
I have **limited** understanding of the recording.

**Band 5**
The speaker's accent is **noticeably different** from what I am used to.
I have to **excessively** concentrate on listening.
I can **barely** understand the recording.

# Appendix B: The Accent Strength and Identification Task

## I. The Background Section

1. How would you describe your English listening skills?
☐ very limited ☐ limited ☐ average ☐ good ☐ excellent

Do you think you are good at telling English accents apart?
☐ not at all ☐ limited ☐ average ☐ good ☐ excellent

2. Is English your mother tongue? If not, please write down your mother tongue(s).
(It is possible that you might have more than one mother tongue if you grew up in a multilingual community)
Yes ☐
No ☐
Other mother tongue(s):

3. Overall, how familiar are you with the following English accents?
Australian English accent
☐ no knowledge ☐ a little familiar ☐ average ☐ familiar ☐ very familiar

Spanish English accent
☐ no knowledge ☐ a little familiar ☐ average ☐ familiar ☐ very familiar

Vietnamese English accent
☐ no knowledge ☐ a little familiar ☐ average ☐ familiar ☐ very familiar

Mandarin English accent
☐ no knowledge ☐ a little familiar ☐ average ☐ familiar ☐ very familiar

## II. The accent judgment section

Listen to each recording clip, decide the accentedness and identify the accent of each clip.
Clip 1:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 2:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 3:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 4:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 5:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 6:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 7:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 8:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 9:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

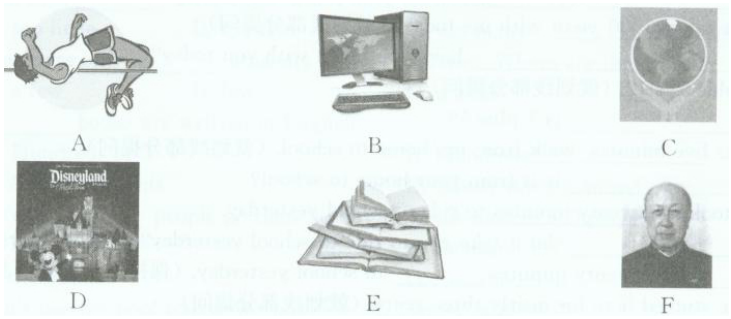
Clip 10:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 11:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

Clip 12:
☐ Band 1 ☐ Band 2 ☐ Band 3 ☐ Band 4 ☐ Band 5
☐Australian ☐Spanish ☐Vietnamese ☐Mandarin

# Appendix C: The Accented English Listening Test

**I. Listen and choose the right picture.**




1. _____ 2 . _____ 3 . _____ 4 . _____ 5 . _____

**II. Listen to the passage and tell whether the following statements are true or false.**

( ) 6. The wife was a large woman and the husband was a small man.

( ) 7. The man worked as a manager in a big company.

( ) 8. The wife gave her weekly money to her husband.

( ) 9. The wife would take all the money from her husband and never give him any money.

( ) 10. The husband was happy to tell his wife that he had won a lot of money.

**III. Listen to the passage and complete the following sentences.**

11. The sea is the _____ of millions of living things.

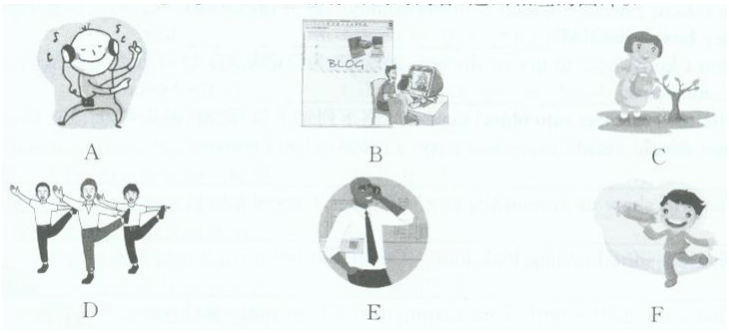
12. These is more life in the sea than on _____ .

13. The animals and plants of sea are very _____ to man as a source of food.

14. Today more and more people are _____ the sea.

15. The land will not be _____ to provide food for everybody.

**IV. Listen and choose the right picture.**




16. _____ 17 . _____ 18 . _____ 19 . _____ 20 . _____

**V. Listen to the passage and tell whether the following statements are true or false.**

( ) 21. Young people smoke, because they think it is cool.

( ) 22. The famous star Jackie Chan also thinks it is cool to smoke.

( ) 23. Yao Ming isn't so cool because he doesn't smoke cigarettes.

( ) 24. Some young people smoke because they see their parents do that.

( ) 25. Exercise is a good way to help us give up smoking.

**VI. Listen to the passage and complete the following sentences.**

26. Jack has _____ hair and blue eyes.

27. Peter and Jack are both on school _____ team.

28. Linda won the women's first _____ of their school.

29. Betty came to Peter's class _____ months ago.

30. Peter _____ here from Europe.

# Appendix D: The Accent Perception Questionnaire

| Item | Statements | Strongly agree | Agree | Slightly agree | Neutral | Slightly object | Object | Strongly object |
|---|---|---|---|---|---|---|---|---|
| **1** | I am familiar with the accent in the test | | | | | | | |
| **2** | I think I can understand the accent in the test | | | | | | | |
| **3** | I hope in future English classes I can hear more of the accent of the test | | | | | | | |
| **4** | I think the accent of the test is similar to the accent of my textbook recordings | | | | | | | |
| **5** | The accent of the test impeded my understanding of the test content | | | | | | | |
| **6** | I find it necessary to develop the ability to understand different English accents | | | | | | | |
| **7** | I think the accent of the test is similar to our English teacher's accent | | | | | | | |

| Item | Statements | Strongly agree | Agree | Slightly agree | Neutral | Slightly object | Object | Strongly object |
|---|---|---|---|---|---|---|---|---|
| **8** | The accent of the test made it hard for me to concentrate | | | | | | | |
| **9** | I like the accent of the test | | | | | | | |
| **10** | If the accent of the test were the accent of our textbook recordings, I would have performed better in the test | | | | | | | |
| **11** | The accent of the test made the test harder than usual | | | | | | | |
| **12** | I hope I can understand different English accents | | | | | | | |
| **13** | If the accent of the test were the accent of our English teacher, I would have performed better in the test | | | | | | | |
| **14** | Due to the accent, the test became easier | | | | | | | |
| **15** | I find the accent of the test very weird | | | | | | | |