To be published in *International Journal of Bilingual Education & Bilingualism* (Routledge)


# The relationship between multilingual raters' language background and their perceptions of accentedness and comprehensibility of second language speech

Natsuko Shintani, Kazuya Saito, & Rie Koizumi

**Abstract**

The purpose of this case study is twofold: it (1) explores multilingual raters' judgements regarding the accentedness and comprehensibility of second language speech and (2) examines how the raters' language backgrounds influence their judgements. In this study, six multilingual Singaporean raters judged the accentedness and comprehensibility of 50 unfamiliar accented speech samples produced by Japanese learners of English with different proficiency levels. In order to investigate rater judgement, the rating scores were subjected to a multifaceted Rasch analysis. A questionnaire and an interview elicited the raters' retrospective reports on their language backgrounds at three time points in their life (when they were 5, 11, and 21 years old). The results suggested that the raters' language backgrounds, notably the proficiency balance between multiple languages in early childhood, are related to their rating judgement.

Keywords: multilingualism; foreign accent, speech perception, rater severity; second language pronunciation

The relationship between multilingual raters' language background and their perceptions of accentedness and comprehensibility of second language speech

## Introduction

Comprehensibility refers to 'the listener's perception of how easy or difficult it is to understand a given speech sample', while accentedness is defined as 'how different a pattern of speech sounds compared to the local variety' (Derwing and Munro 2009, 478). There is evidence that comprehensibility is closely related to, but independent of, accentedness (Derwing and Munro 2009). The comprehensibility and accentedness of second language speech is apparently determined by the linguistic, particularly phonetic, quality of the speech (Munro and Derwing 2001; Trofimovich and Baker 2006). However, it might also be influenced by listener factors such as first language and familiarity with accented speech.

The role of listener's language background on their speech evaluation is an increasing interest of researchers (Kennedy and Trofimovich 2008; Carey, Mannell, and Dunn 2011; Winke, Gass, and Myford 2013). Such studies have shown mixed evidence. A number of studies investigated the influence of listeners' first language (L1) on comprehensibility judgement (Carey, et al. 2011; Winke et al. 2013). These studies have shown that listeners tend to evaluate L2 speech more leniently when they recognize that they share the same L1 as the speaker. For example, Major, Fitzmaurice, Bunta, and Balasubramanian (2002) showed that non-native speakers often understand an L2 speaker from their own L1 background more easily than from a different background. Winke et al. (2013) investigated 107 raters who were L2 speakers of Spanish, Chinese, or Korean and evaluated the speech samples of three native-speaker groups (24 speakers each) in these languages. The results of a multifaceted Rasch analysis indicated that L2 Spanish raters were significantly more lenient towards L1 Spanish test takers, as were L2 Chinese raters towards L1 Chinese test takers.

The subsequent examination of raters' self-evaluation of their rating process indicated that the raters' awareness of accents and familiarity with specific accents influenced the rating process (Winke and Gass 2013). Winke and Gass explained the reasons as the phonological similarity between the L1 and L2 and the possibility that the repeated exposure to accented speech had increased the speed of processing the words and sentences (Weil 2001).

There are also a number of studies examining listeners' general familiarity with foreign accents (Isaacs and Thomson 2013; Kennedy and Trofimovich 2008). These studies have shown that native speakers with previous exposure to non-native speech are likely to rate accentedness and comprehensibility more leniently than those without such exposure. Some studies have shown more specifically that familiarity with a particular accent increases the comprehensibility of the accented speech (Gass and Varonis 1984; Major et al. 2002).

However, other studies suggested a limited influence from the listener's language background. Munro, Derwing and Morton (2006) investigated listeners with native Cantonese, Japanese, Mandarin and English backgrounds, evaluating the same set of foreign-accented English speech by native speakers of Cantonese, Japanese, Polish and Spanish. Regardless of native language background, there were moderate to high correlations between the intelligibility scores and the comprehensibility and accentedness ratings. MacKay, Flege and Imai (2006) examined the ratings of accented speech between listeners who were native speakers (NS) and non-native speakers (NNS); however, finding that these listeners demonstrated similar severity patterns (see also Kim, 2009).

As Munro (2008) argued, the relationship between listeners' language background and L2 comprehensibility judgements could be considered as a complex, multifaceted phenomenon characterized by a range of factors throughout a life —e.g., first language background, familiarity with particular foreign accents, own foreign language experience.

This line of thought echoes Bradlow and Bent's (2008) notion of "perceptual adaptation" which states that humans maintain the capacity to flexibly, dynamically and constantly adjust their mental retransitions according to their listening experience. Whereas inexperienced L2 learners may prefer their own L1-accented speech due to the lack of exposure to target language (e.g., Major et al., 2002), experienced L2 learners tend to have much experience with target language and approximate NSs' speech judgement patterns (e.g., Munro et al., 2006). At the same time, NSs themselves also likely have different perceptual representations, as every individual must have different listening experience with various interlocutors in diverse social settings (Norris, McQueen, and Cutler, 2003). In essence, the previous literature presented here suggests that the intricate interaction between listener backgrounds and behaviours needs to be more carefully/closely examined via not only "quantitative" but also "qualitative" analyses from multiple angles. To our knowledge, however, such in-depth analyses are rare in the field of L2 speech research.

One study investigated yet another aspect of the listeners' language background, i.e., their ability to use multiple languages (i.e. monolingual versus multilingual listeners). Saito and Shintani (2016) compared how monolingual L1 speakers in Canada and multilingual L1 speakers in Singapore judged the comprehensibility of unfamiliar L2 speech. In these studies, two groups of raters, 10 monolingual native speakers of Canadian English and 10 multilingual speakers of Singapore English (SE), had to judge the L2 comprehensibility of Japanese English speakers having a range of proficiency levels. The results showed that on comparing the scores, the Singaporean raters showed significantly greater leniency in their judgements than the Canadians for comprehensibility. While both groups of raters used the same processing strategy to make comprehensibility judgements (exclusively drawing on segmental accuracy), the Singaporean raters used more lexicogrammatical information when evaluating comprehensibility than the Canadians. These findings have further raised the question as to what attributed these multilingual raters' lenient judgements.

One possible factor is the listeners' other fluent language(s). Eight of the ten participants in Saito and Shintani were fluent in Mandarin in addition to English. The similarity in the phonetic patterns of Mandarin to Japanese, such as lack of vowel reduction (Major et al. 2002) might have led to the listeners' lenient judgements. Another possible reason is, as Saito and Shintani suggest, Singaporeans' rich exposure to English varieties with substantial differences in acoustic realization among ethnic groups (Tan 2012). Such constant need to accommodate a variety of accented English results in Singaporeans being able to recognize the ethnicity of a speaker from a short speech sample (e.g., Deterding and Poedjosoedarmo 2000). However, little – if any – research scrutinized multilingual listeners' language background in relation to their judgement of comprehensibility and accentedness.

To explore this issue, the current case study revisited the Singaporean multilingual raters' responses to L2 speech in the study by Saito and Shintani (2016), focusing on those who shared the same ethnic background (i.e. Chinese). By triangulating not only the quantitative analysis of their rating results but also the qualitative analysis of their retrospective reports, we aimed to examine in detail the complex relationship between rater types (lenient vs. strict), rater domains (comprehensibility and accentedness) and rater background (their multilingual experience at the age of 5, 11, and 21).

**Multilingual context in Singapore**

In this study, 'multilingualism' refers to a social situation in which more than two languages are used commonly and officially. Singapore has four official languages: English, Mandarin, Tamil, and Malay. Although English serves as the lingua franca in most formal contexts (e.g. schools, work, and the media), all Singaporeans are required to learn any of Mandarin, Tamil, or Malay in earlier classes since high-stakes examinations include the

ethnic languages. In addition to the official languages, many other languages and dialects (e.g. Hokkein) are used informally or by local communities. According to Cavallaro and Chin (2014), English is not the dominant language at home for many Singaporeans. Investigating Singaporean university students' language use, Aman et al. (2007) reported that 71% of the students had habitual bilingual practices, using both English and their Mother Tongue, as their dominant home language; 7% saying they use only English; and 19% saying only their Mother Tongue is used at home.

Singapore also has a unique demography of English varieties. The spoken English in Singapore is called Singapore English (SE). Although SE historically originated from British English, the prevalence of rhoticity (a trait of American English) in native SE speakers and distaste for the intrusive-*r* in non-rhotic accents (Tan and Gupta 1992; Tan 2012) are increasing. Further, SE supports a range of unique linguistic phenomena in terms of phonology, lexes, and morphosyntax. SE is used very commonly in an informal context in Singaporeans' daily life. In addition to SE, Singaporeans are regularly exposed to various non-native English accents originating from the Philippines, India, Hong Kong, China, Indonesia, and Malaysia (Lalwani, May, and Kuah 2005) due to the presence of a large number of immigrants (386,000 from Malaysia; 175,200 from China, Hong Kong, and Macau; 123,500 from South Asia; 54,400 from Indonesia; and 90,100 from other Asian countries in 2010) (Yeoh and Weiqiang 2012) and the increasing number of foreign workers (1,368,200 in 2015, as reported by the Ministry of Manpower [2015]). Apart from the English used by immigrants and foreign workers, Singaporeans are regularly exposed to North American and British English through the media. Specifically, the younger generation is generally influenced by such media through the Internet (Tan and Tan 2008).

As suggested by Clyne, Rossi Hunt, and Isaakidis (2004), it is not easy to quantify the multilingual speakers' dynamic and diverse use of language, which is strongly influenced by perceived communicative needs. Individuals' language profiles are established through changes that take place over time in terms of language proficiency and, consequently, language dominance in a multilingual repertoire (Jessner 2008). The current study, thus, focused on individual Singaporeans' language experiencing a retrospective self-report. It also explored how such backgrounds differed among the individuals who judged the accentedness and comprehensibility of unfamiliar L2 speech. Based on these considerations, this study addresses the following research question: Do multilingual raters' language backgrounds relate to their rating of the accentedness and comprehensibility of unfamiliar accented speech? Different from previous studies on L2 listener factors, the current study was the first attempt to approach this topic via incorporating the quantitative and qualitative analyses of multiple data resources—rating scores, linguistic analyses, questionnaires and interviews.

## Methods

### *Participants*

Six students were selected from the participants in Saito and Shintani (2015). In Saito and Shintani, the participants were recruited from three universities in Singapore according to the following criteria: 1) one or both of their parents spoke English as a primary language; 2) they had no previous experience learning Japanese or exposure to Japanese-accented English and 3) they had no experience in English-language teaching. The ten students who participated in Saito and Shintani (2015) were further screened for the current study with the condition of being ethnic Chinese. The decision was made to control the variations in familiarity with languages other than English. Ethnic Chinese was chosen because it was the major ethnic group in Singapore (76.2% in 2014, according to the Ministry of Home Affairs [2014]).

All of the six participants had attended public schools before joining the university, receiving English-medium education throughout; further, they had attended three to four hours a week of Mandarin classes, at least in primary and secondary schools. While participating in this study, five raters lived in a university dormitory, whereas one (Rater 3) lived with her family.

***Data analyses***

To examine the link between rater backgrounds and behaviours, our study consisted of three different data analysis stages: (a) L2 speech judgement analyses; (b) questionnaire analyses; and (c) interviews.

***1. Accentedness and comprehensibility analyses***

At this quantitative stage, we first asked six raters to rate 50 speech samples (produced by Japanese learners) for accentedness and comprehensibility on a 9-point scale. Subsequently, Rasch analyses were conducted to identify the relatively strict and lenient raters.

*Materials*

The test used the oral stimuli developed by Saito (2011), comprising 50 speech samples made by Japanese English language learners having a wide range of proficiency levels. The speech data were developed based on a corpus of audio recordings of 200 Japanese learners living in Canada (Saito 2011). Each audio stimulus, approximately 10 seconds long, was extracted from a 30-second speech production elicited using three picture description tasks (see the study by Saito, Trofimovich and Isaacs 2015).

*Procedure*

The method used for rating accentedness and comprehensibility followed that of previous research (Derwing and Munro 1997; Trofimovich and Isaacs 2012). The sessions took place individually in quiet rooms at the chosen universities in Singapore. On familiarizing themselves with the picture prompts, the raters were presented with the speech samples in a randomized order using the Praat software (Boersma and Weenik 2012). They were asked to listen to each picture description only once, rate accentedness on a 9-point numerical scale (1 = no accent and 9 = heavily accented), and subsequently rate comprehensibility using the scale of 1 = easy to understand and 9 = hard to understand. The raters were reminded to use the entire scale in judgement. The judged scores for each speech sample were digitally recorded using Praat and later transferred to an Excel spreadsheet. A five-minute break was provided halfway through the sessions to avoid fatigue. The entire session lasted approximately 90 minutes.

*Rasch analyses*

In order to examine the reliability and the differences in rating severity and consistency, the accentedness and comprehensibility judgement scores provided by the six raters were subjected to a multifaceted Rasch analysis using Facets (Linacre 2014) for each score type. This analysis was used since it can estimate rater severity more precisely than the analysis using raw ratings, while considering differences in task difficulty and speakers' ability (Barkaoui, 2013). The unit of the logit scale (in which the parameters for this model were estimated) was set as 0 for the task difficulty and speakers' ability. As such, the analyses here were expected to allow us to identify which raters would be the most "strict" and "lenient" from a statistical point of view.

***2. Questionnaire analyses***

To further provide the retrospective language experiences and self-report language proficiency of the strict/lenient raters (identified via Rasch analyses), the questionnaire data were used. Their answers to the questionnaire items were chronologically organized by the researcher to present the individual raters' language experience and perceived proficiency (see Appendix B). To analyse the linguistic profiles of the raters, three time points were

chosen (when the participants were 5, 11, and 21 years old) for investigation based on the following considerations: The ages from zero to five were considered the 'preschool' period, as children in Singapore graduate preschool at age five, by which their language development is more or less complete. Supposedly, home languages greatly affect language development during this period. Age 11 signified the completion of primary school and the period when the participants first experienced a new language domain (i.e. school), which had presumably influenced their language use significantly. Age 21 was chosen to ascertain their present status, as all the participants were 21 or 22 years old when this study was conducted.

The questionnaire consisted of three sections, Section A: family languages, Section B: the use and proficiency of languages, and Section C: exposure to English varieties (see Appendix A for the entire questionnaire). Sections B and C consisted of three subsections for ages 5, 11, and 21. The participants were requested to provide information about their language exposure with their interlocutors or sources in as much detail as possible. We instructed the participants to consider Chinese dialects (e.g. Hokkein) as 'languages' rather than 'dialects' as they were not intelligible to those who spoke another variety of Chinese (e.g. Mandarin). The self-evaluation of their language proficiency involved four categories (1 = beginner, 2 = low intermediate, 3 = intermediate, 4 = advanced, and 5 = native). One of the researchers distributed the questionnaire through email to all the participants who had completed the judgement tests and collected it within a week, again through email.

***3. Interview***

Two of the participants, one most lenient and one least lenient based on the accentedness and comprehensibility judgement test, were selected for further investigation.

*Procedure*

One of the researchers conducted a follow-up individual interview with the two raters using an online videoconferencing programme, *Skype*, within one week (two to seven days) of the questionnaire's completion. The interview focused on eliciting retrospective experiences and anecdotal events that might have influenced the participants' use of languages or English varieties and their language proficiency. The interview included the following questions:

- Could you tell me more about your use of [English/Mandarin]?
- When did you use English with your [mother/father/brother]?
- How often did you watch TV in [English/Mandarin]? / With whom did you watch the programme?
- Did you experience any difficulty in understanding your [grandmother/friends] when [she/they] spoke to you in [Hokkein/Mandarin]?
- Do you think your [English/Mandarin] was better than that of your friends? What made you think so?

The interview was audio recorded on the interviewer's computer for analysis.

*Analyses*

The interview data for the two raters was audio-recorded on the interviewer's computer and transcribed by the researcher. The raters' utterances were coded in terms of the three time points set for this study (i.e., early childhood, school and current). The data were further coded in terms of the three focuses of the questionnaire (i.e., language use, language proficiency and exposure to English varieties). One utterance could be coded as two categories. For example, an utterance, 'I was quite close to my grandmother; so, I had no problem with conversing with her', was coded as 'language use' as well as 'language proficiency'. The transcripts were coded independently by two raters (one of the researchers and a graduate research assistant). Disagreements between coders were resolved by discussion. The coded data for the two raters were analysed by the researchers to identify

contrasting statements between the two raters and were used to explain the quantitative results.

## Findings

This section first reports the results of the multifaceted Rasch analyses to examine whether the six raters judged the L2 speech differently. As such, we aimed to statistically analyse and identify the relatively strict vs. lenient raters. Next, we investigated the questionnaire data to expound the patterns in terms of the language background of such strict/lenient raters. Finally, the interview data of two raters, the strictest and the most lenient, were used to confirm and further explore the quantitative results.

### *Results for the Multifaceted Rasch analyses*

*Accentedness scores.*

Table 1 shows the calibration report for the raters. The second column shows the means of observed raw scores given by the raters (larger values indicate more lenient judgement). The third column is the logit measure for the raters ($M = 0.00$, $SD = 0.41$), with the most lenient rater at the top and the most severe rater at the bottom. The standard errors of the logit measures (column 4) were quite low, 0.06 or 0.07, indicating high precision in measurement.

Columns five and six are the 'quality control' fit statistics, which show the goodness of fit of the data to the Rasch model in residual mean squares. Infit is an information-weighted fit statistic and insensitive to outliers. Outfit is an unweighted fit statistic, sensitive to outliers and is 'dominated by unexpected outlying, off-target, low information responses' (Linacre and Wright 1994). The suggested range for an acceptable fit of infit and outfit mean squares is 0.5 to 1.5; raters below that range are considered overfit and unexpectedly consistent; raters above this range are considered underfit and unexpectedly inconsistent (Linacre 2013). According to this range, Rater 5 overfit the model with an infit mean square of 0.43. In many cases, this rater tended to give the same score across all or most categories, indicating a possible halo effect (Engelhard and Stone 1998). Although Rater 3 underfit the model with an infit mean square of 1.69, the value was within 1.5 to 2.0, which was suggested as 'unproductive but not degrading' (Linacre 2013, 266)[1]. Columns seven and eight show a 95% confidence interval (CI) of rater severity measure defined by logit ±1.96 × standard error.

The rater separation ratio was 6.09, meaning that the six raters could be divided into six groups in terms of severity. The reliability of separation was .99, and the fixed (all-same) chi-square was significant at the $p = .01$ level, indicating that the raters were not at the same severity level. Although it is usually considered undesirable to have reliable separation between raters (Lunz and Linacre 1998) because we were looking for comparable ratings of the test candidates, in this study the raters clearly differed from each other.

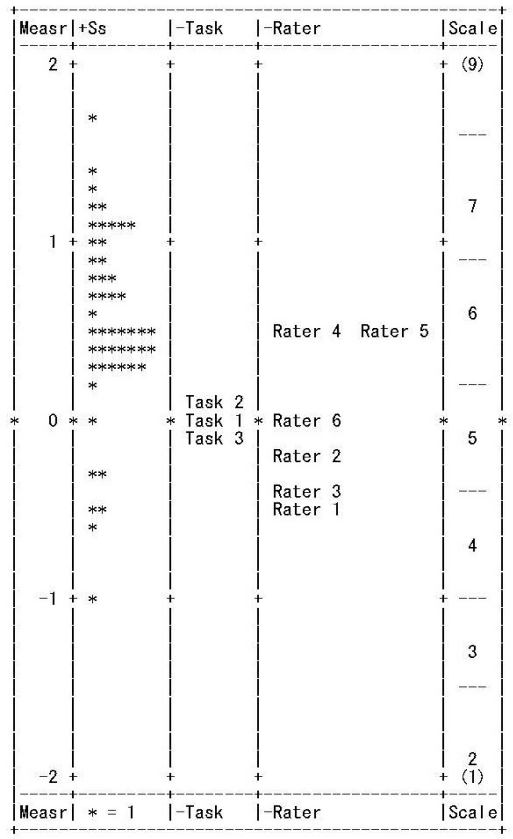
As the 95% confidence interval indicates, significant differences between raters can be identified as: Raters 4 and 5 > Raters 1, 2, 3 and 6; Rater 6 > Raters 1 and 3; Rater 2 > Raters 1. This indicates that there were the raters were divided into three severity groups with from the most lenient: Raters 4 and 5 > Raters 2 and 6 > Raters 1 and 3.

Table 1. The Facets calibration report for six raters in accentedness judgements.

| Rater | Observed average | Logit | Standard Error | Infit MnSq | Outfit MnSq | 95% CI | |
|---|---|---|---|---|---|---|---|
| | | | | | | Low | High |
| Rater 5 | 5.13 | 0.54 | 0.06 | 0.43 | 0.43 | 0.42 | 0.66 |
| Rater 4 | 5.15 | 0.53 | 0.07 | 1.31 | 1.29 | 0.39 | 0.67 |
| Rater 6 | 6.01 | -0.03 | 0.07 | 0.57 | 0.58 | -0.17 | 0.11 |
| Rater 2 | 6.25 | -0.18 | 0.07 | 0.58 | 0.58 | -0.32 | -0.04 |
| Rater 3 | 6.51 | -0.35 | 0.07 | 1.69 | 1.69 | -0.49 | -0.21 |
| Rater 1 | 6.75 | -0.51 | 0.07 | 1.45 | 1.46 | -0.65 | -0.37 |

Note. MnSq = Mean-square. Separation: 6.09. Reliability: .97. Fixed (all same) chi-square: 229.0, $df = 5$, $p < .01$.

A Wright map of the speakers' ability, task difficulty and rater severity is given in Figure 1. The first column shows the logit scales, and the second, third, and fourth column shows speakers' accentedness (with higher values meaning higher accentedness), task difficulty (with higher values meaning more difficult tasks), and rater severity (with higher values meaning more lenient raters). The last column shows the structure of the accentedness rating scale. Figure 1 indicates that the speakers' accentedness varied widely (logit = -1.01 to 1.73; $M = 0.54$, $SD = 0.53$) and that difficulty of the three tasks differed little across the tasks (logit = -0.10 to 0.13; $M = 0.00$, $SD = 0.10$); this was in line with our intentions in planning the research design.




Note: Speakers and tasks all fit the model except for nine speakers, but eight had the infit mean square of lower than 2.00.

Figure 1. Wright map of speech production level, task difficulty and the rater severity for accentedness scale.

*Comprehensibility scores.*

Table 2 shows the calibration report for the raters' judgement of comprehensibility. The standard errors were low, ranging from .05 to .07, indicating high precision in measurement. The six raters' rating scale data fit into the range of .50 to 1.50. Thus, all raters were considered to show reasonable consistency in their rating.

The rater separation ratio was 6.35, meaning that there were six groups of raters in terms of severity (i.e. the six raters were different). The reliability of separation was .98, and the fixed (all-same) chi-square was significant at the $p = .01$ level, indicating that raters were not at the same severity level.

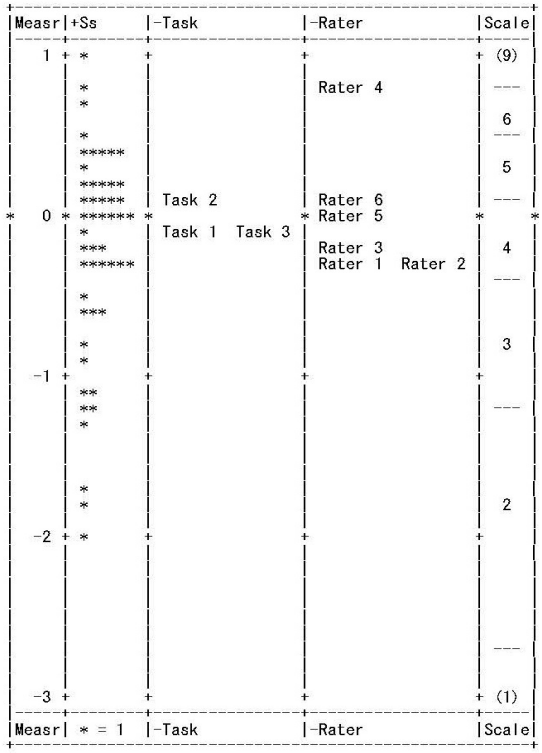
Table 2. Facets calibration report for 6 raters in comprehension judgements.

| Rater | Observed average | Logit | Standard Error | Infit MnSq | Outfit MnSq | 95% CI Low | 95% CI High |
|---|---|---|---|---|---|---|---|
| Rater 4 | 2.87 | 0.78 | 0.07 | 0.78 | 0.75 | 0.64 | 0.92 |
| Rater 6 | 3.96 | 0.08 | 0.06 | 1.16 | 1.08 | -0.04 | 0.20 |
| Rater 5 | 4.11 | 0.00 | 0.06 | 1.20 | 1.27 | -0.12 | 0.12 |
| Rater 3 | 4.63 | -0.24 | 0.05 | 1.34 | 1.31 | -0.34 | -0.14 |
| Rater 2 | 4.77 | -0.30 | 0.05 | 0.68 | 0.74 | -0.40 | -0.20 |
| Rater 1 | 4.81 | -0.32 | 0.05 | 0.84 | 0.90 | -0.42 | -0.22 |

Note: Separation: 6.35. Reliability: .98. Fixed (all same) chi-square: 191.7, $df = 5$, significance: $p < .01$.

The 95% confidence interval indicates that significant differences were identified among the following: Raters 4, 5 and 6 > Raters 1, 2 and 3; Rater 4 > Raters 5 and 6. This result indicates that the raters could be categorised into three groups in terms of their severity with from the most lenient: Rater 4 > Raters 5 and 6 > Raters 1, 2 and 3.

Figure 2 shows a Wright map of the speakers' ability, task difficulty and rater severity for the comprehension scale. Figure 2 indicates that the speakers' comprehensibility (the second column) varied widely (logit = -2.02 to 0.97; $M = -0.22$, $SD = 0.66$) and that difficulty of the three tasks (the third column) differed little across the tasks (logit = -0.08 to 0.14; $M = 0.00$, $SD = 0.10$); this was in line with our intention in planning the research design.




Note: Speakers and tasks all fit the model except for five speakers, but four had an infit mean square of lower than 2.00.

Figure 2. Wright map of speakers' ability, task difficulty and the rater severity for comprehensibility scale.

*Summary for the Rasch analysis results.*
The above analyses demonstrated that the six raters showed statistically significant differences in both their accentedness and comprehensibility judgements. The analysis identified three levels of severity in accentedness and comprehensibility judgements (Figure 3). For accentedness, Raters 1 and 3 were the strictest; Raters 4 and 5 were the most lenient; and Raters 2 and 6 fell in the middle. For comprehensibility, Raters 1, 2 and 3 were the strictest; Rater 4 was the most lenient; and Raters 5 and 6 fell in the middle.

| | Stricter ⟷ | | More lenient |
|---|---|---|---|
| Accentedness | R1 and R3 | R2 and R6 | R4 and R5 |
| Comprehensibility | R1, R2 and R3 | R5 and R6 | R4 |

Figure 3. Summary of raters' severity groups.

Figure 3 indicates that the six raters judged the L2 speech differently for both comprehensibility and accentedness. Adding to the findings of Saito and Shintani (2016), suggesting that multilingual L1 speakers in Singapore judged the comprehensibility of an unfamiliar L2-accented speech more leniently than monolingual L1 speakers in Canada, the results show evidence that within the same context (i.e., Singaporean), raters varied in their judgement of comprehensibility and accentedness of L2 speech.

Figure 3 also suggest that there is tendency for an association between accentedness and comprehension judgement in terms of the raters' severity (i.e., raters graded accented severely also tended to grade comprehensibility severely). However, the average scores of both measurements (Table 1 and Table 2) indicate that the raters were inclined to judge accentedness more severely than comprehensibility. The findings agree with the argument by Derwing and Munro (2009) that these two elements of L2 speech perception are different but closely related. The results also indicate that the raters tended to evaluate some of the L2 speech as accented but comprehensible, supporting the argument by Derwing and Munro.

For the subsequent qualitative analyses (questionnaire, interviews), we drew on the results of the Rasch analyses to focus on Rater 3 and Rater 4 as the most "strict" and "lenient" raters, respectively. Both of them were categorized as either "strict" or "lenient" regardless of the accentedness and comprehensibility judgements. In what follows, we provided the results of the qualitative analyses (questionnaires, interviews) of the two raters.

***Rater judgement and language background***

In this section, the questionnaire data were used to compare the language backgrounds of the strict rater (Rater 3) and the lenient rater (Rater 4). For the purpose of comparison, we also looked at the results of the other raters (Raters 1, 2, 5, 6). Finally, the focused interview data of the most severe rater (Rater 3) and lenient rater (Rater 4) were used to add detailed information. We will examine the results in terms of a) language exposure, b) language proficiency and c) English varieties.

*Language exposure*

Table 3 summarises the questionnaire results (see Appendix B for the entire questionnaire results). The most noticeable difference among raters is the ratio of the exposure to different languages in early childhood. Three raters who were relatively lenient in evaluation reported that their English exposure was 50% or less in early childhood (Raters 4, 5 and 6), while two of the strict raters reported that English was the predominant language in their childhood (90% and 85% for Raters 3 and 1, respectively). These differences diminished once the raters entered their primary schools and spent longer periods at school and with friends.

Table 3. The summary of the raters' language experience.

| Rater | | Language experience elements | Preschool (age 5) | School (age 11) | University (age 21) |
|---|---|---|---|---|---|
| R1 | Exposure | Number of languages | 2 | 2 | 2 |
| | | English exposure (%) | 85 | 90 | 90 |
| | | Number of English varieties | 3 | 3 | 5 |
| | Proficiency | English (1-5) | 5 | 5 | 5 |
| | | Other languages (1-5) | 2 (M), 1 (T) | 2 (M) | 3 (M) |
| R2 | Exposure | Number of languages | 3 | 3 | 2 |
| | | English exposure (%) | 60 | 70 | 80 |
| | | Number of English varieties | 3 | 3 | 5 |
| | Proficiency | English (1-5) | 5 | 5 | 5 |
| | | Other languages (1-5) | 3 (M), 1 (H) | 4 (M) | 4 (M) |
| R3 | Exposure | Number of languages | 3 | 2 | 2 |
| | | English exposure (%) | 90 | 80 | 90 |
| | | Number of English varieties | 3 | 3 | 5 |
| | Proficiency | English (1-5) | 5 | 5 | 5 |
| | | Other languages (1-5) | 1 (M) | 3 (M) | 3 (M) |
| R4 | Exposure | Number of languages | 4 | 3 | 3 |
| | | English exposure (%) | 50 | 70 | 80 |
| | | Number of English varieties | 4 | 3 | 4 |
| | Proficiency | English (1-5) | 3 | 5 | 5 |
| | | Other languages (1-5) | 4 (M), 3 (H) | 5 (M), 3 (H) | 5 (M), 1 (H) |
| R5 | Exposure | Number of languages | 3 | 2 | 2 |
| | | English exposure (%) | 50 | 70 | 70 |
| | | Number of English varieties | 3 | 4 | 5 |
| | Proficiency | English (1-5) | 5 | 5 | 5 |
| | | Other languages (1-5) | 4 (M), 1 (H) | 4 (M) | 4 (M) |
| R6 | Exposure | Number of languages | 3 | 3 | 2 |
| | | English exposure (%) | 40 | 80 | 80 |
| | | Number of English varieties | 3 | 3 | 4 |
| | Proficiency | English (1-5) | 3 | 5 | 5 |
| | | Other languages (1-5) | 4 (M), 1 (H) | 5 (M), 1 (H) | 5 (M), 1 (H) |

Proficiency: 1 = beginner, 2 = low intermediate, 3 = intermediate, 4 = advanced, 5 = native level.
Other languages: M = Mandarin, T = Teochew, H = Hokkein.

Table 4 summarizes the most lenient (Rater 4) and the strictest (Rater 3) raters' early childhood language exposures and the resources of the languages at home and in preschool. The members specified with an asterisk are those who lived with the rater. Rater 3 reported that she used English 90% of the time with her parents and her home helper. She was exposed to Mandarin and Teochew 10% but she did not live with any of the speakers of these languages. At the preschool where Rater 3 was enrolled when she was three, she used English 80% of the time. Rater 4, on the other hand, reported more mixed language use at home. She

used English 50% with her father, brother and home helper and other languages (Mandarin and Hokkein) with her mother, brother and grandmother. All of these members lived with her. Her exposure to English (70%) and Mandarin (30%) at preschool was similar to that of Rater 3. Both Raters 3 and 4 watched TV in English, but only Rater 4 watched Mandarin programmes.

Table 4: The resources of language exposure in the preschool phase (at 5 years of age)

| | | Home | Preschool |
|---|---|---|---|
| R3 | English | 90%<br>Mother*<br>Father*<br>Home helper*<br>Father's family<br>TV | 80%<br>Teacher<br>Friends |
| | Other languages | 10%<br>Mother's family (Mandarin)<br>Cousins (Mandarin)<br>Grandmother (Teochew) | 20%<br>Mandarin teacher (Mandarin)<br>Friends (Mandarin) |
| R4 | English | 50%<br>Father*<br>Brother*<br>Home helper*<br>TV | 70%<br>Teacher<br>Friends |
| | Other languages | 50%<br>Mother* (Mandarin)<br>Brother*<br>Grandmother* (Hokkein)<br>Relatives (Mandarin)<br>TV | 30%<br>Mandarin teacher (Mandarin)<br>Friends (Mandarin)<br>Friends (Malay) |

*A person who lived with the rater.

The two raters' retrospective reports about their language exposure showed very different demography for their language use in their early childhood, reflecting the above tendency. Rater 4 had family members who spoke English (her father) and other languages (her mother, brother and grandmother) when she was age five. She recalled that although she used to talk to her brother in Mandarin when she was young, after entering primary school, they started to use English, particularly when they watched American cartoon TV programmes. She recalled her experience as follows.

> My father talked to me in English and my mother talked to me in Mandarin. But they talked to each other in Mandarin. We lived with my grandmother, who spoke Hokkein most of the time. I understood her and talked to her in Hokkein. I was quite close to my grandma when I was young. When we met with relatives, we talked in Chinese all the time. (Mandarin?) Yes. We had a helper from Indonesia. She spoke to me in English. (Did she speak in accented English?) Yeah, very much. With my brother, we always spoke in English. When we watched TV with my parents, it was usually in Chinese (Mandarin), but cartoons were often in English-American (English). My father spoke to me in English because English is important for me.

The exposure to particular languages seemed to inter-relate with the rater's closeness with particular family members. Rater 4's comments in the interview suggested that she was very

close to her grandmother, with whom she spent most of her time when her parents were working:

> Because both of my parents were working, I spent most of my time with my grandmother. (What language did you use?) I talked to her in Hokkein. Actually, I was quite close to my grandmother; so, I had no problem with conversing with her. She told me lots of stories, and I talked to her about many things.

Rater 3 reported a contrasting language experience from Rater 4. All her family members at home (i.e., her mother, father and home helper) predominantly spoke English. The only opportunities to hear other languages were when her family visited her relatives (her grandparents, cousins and aunt), which she reported as 10% of her childhood life. She stated that although some of her relatives spoke other languages, she was not close to them. She contacted her relatives approximately once a week and communicated with them using Mandarin. Her preschool teachers and friends mostly spoke English, with some of them speaking Mandarin. According to Rater 3, 'I speak with my friends in English most of the time because most of my friends speak English'. This seems to have resulted in her limited Mandarin proficiency at age five. Her early childhood English proficiency was graded as 5 ("native level"), while Mandarin was graded as 1.

Contrasting to Rater 4, Rater 3 had limited interaction with her grandmother, who did not live with her. Hence, she had only limited receptive Teochew knowledge. She had receptive knowledge but did not seem to have sufficient skills to communicate with her effectively. This seemingly prevented her from establishing a close relationship with her grandmother.

> My grandmother spoke Teochew only. I think I picked up some words and now understand basic things. I wasn't very close to my grandmother. I think this is because I didn't speak her language. I understood what she said, but I didn't talk much with her.

The family language influenced the language exposure in early childhood. Rater 4 had an English-speaking father, a Mandarin-speaking mother, an English- and Mandarin-speaking brother and a Hokkein-speaking grandmother at home, while Rater 3 had a completely English-speaking family (her parents) and home helper. The close relationship of Rater 4 with her grandmother enabled her to develop communicative skills in Hokkein, although her grandmother was the only source of language input.

The above analysis shows the complex nature of language use in multilingual families. Although the raters reported the ratio of their language use at home, the dynamic and diverse use of languages with unequal weighting suggests that quantifying language exposure is not easy (Clyne et al. 2004). The exposed languages seem to change over time as the communicative needs and their proficiency changes (Jessner 2008). We will now move on to the strict and lenient raters' perceived language proficiency.

*Language proficiency*

The questionnaire data (see Table 3) shows that the six raters' self-evaluated language proficiency reflects their language exposure. Raters 4 and 6 (both were lenient in judgement) perceived their early childhood proficiency in a non-English language (Mandarin) to be higher than that in English, although their English did improve to the same level as Mandarin during their school years (both rated their Mandarin and English proficiencies as 'native level'). The other four raters had better command of English than Mandarin during early childhood, which continued up to the last time point (21 years). Raters 1 and 3, who were the strictest in the L2 speech judgement, reported that their early childhood proficiency in English (rated 5) was much higher than that in Mandarin (rated 3 and 1 respectively).

However, after their experiences in English-medium schools (primary, secondary, junior college and university), all the participants stated that they were the most proficient in

English. Two of them (Raters 4 and 6) stated that their proficiency in Mandarin was equivalent to that in English. The changes in perceived proficiency during their school periods reflect the exposure to English at school (see Table 3). Such rapid improvement suggests the strong impact of Singapore's educational system, which emphasises the importance of these two languages in academic success. As Aman et al. (2007) report, university students in Singapore perceive that strong proficiency in both English and mother tongue (i.e., Mandarin), with slightly greater importance placed on the former, was essential to achieve overall academic success.

Table 5 summarises the two raters' (strict or lenient) language proficiency at three time points. In early childhood, Rater 3 developed skills mainly in English and to a much lesser extent in other languages. She evaluated her English skill as 5 (native) throughout the three time points, but graded herself as only 1 for Mandarin at age five, which increased as she grew up but levelled off at 3. Rater 4, on the other hand, acquired relatively balanced knowledge in the three languages she was exposed to (English, Mandarin and Hokkein). She graded her English proficiency as 3 at age five, which was lower than her Mandarin (4) and equivalent to her Hokkein (3) proficiencies. At age 12, she graded her English and Mandarin skills equally as 5. She reported her Hokkein proficiency as 3 at age 12 but it fell to 1 at age 21.

Table 5 Differences in the comprehensibility judgement and language backgrounds of Raters 3 and 4.

| | Rater 3 | Rater 4 |
|---|---|---|
| Observed average score | 4.63 (the strictest) | 2.87 (the most lenient) |
| Proficiency at age 5 | English (5) > Mandarin (1) | Mandarin (4) > English (3) = Hokkein (3) |
| Proficiency at age 12 | English (5) > Mandarin (3) | English (5) = Mandarin (5) > Hokkein (3) |
| Proficiency at age 21 | English (5) > Mandarin (3) | English (5) = Mandarin (5) > Hokkein (1) |

Note: "Mandarin (4) > English (3)" means the self-reported proficiency for Mandarin was '4', which is greater than that of English (graded as 3).

Rater 4 recalled in the interview, 'I was pretty much good at all three, but Mandarin was better. I think it was because I used Mandarin more, like talking with my mother or brother, watching TV and also at preschool'. She used English when communicating with her father, who talked to her in English for educational purposes. However, when asked whether she had any trouble with understanding English, she answered, 'I don't think so, but I just prefer to speak in Chinese than English'. Rater 4's comments indicate that she had developed communicative ability in both English and Mandarin, although she considered herself better in Mandarin than in English. Although she had better command of Mandarin over English, Rater 4 seemed to switch languages with ease depending on the interlocutors. She reported that she had some friends who preferred to speak in English or Mandarin but that she accommodated their friends' preferences when communicating with them. Rater 4 stated, 'It (the language use) really depends. When the person speaks in Mandarin, I speak Mandarin; sometimes, the conversation became mixed up'.

Rater 3, on the other hand, indicated clearer strength in English than other languages in her early childhood as well as her later life. She seemed to have developed her Mandarin knowledge at primary school; however, she confessed that English has been her strongest language all her life.

(In early childhood) I understood my friends or relatives when they spoke Mandarin to me, but I didn't speak much . . . only some words.

(At primary school) my Mandarin was worse than that of other students. My family is an English-speaking one, so we hardly communicate in Mandarin. I am pretty sure that this resulted in me being terrible in Mandarin. Yes, I feel the same now, because I have a basic ability for Mandarin now; I mainly use it to handle common everyday events only.

Rater 3 continued with her reasons for her limited proficiency:

I had difficulties in the language as I needed to study mainly by rote learning at school and it was not naturally easy for me.

(In the Mandarin class in primary school) I couldn't pronounce well. Sometimes, the teacher laughed at me in front of others. So I was quiet . . . I didn't want to make eye contact with my teacher (laugh). I felt I wasn't capable. But I was good in English. I got good feedback from the teacher. (In the Mandarin class) I just felt out of place.

The contrasting statements of the two raters suggest a clear difference in early childhood multilingual proficiency. Acknowledging the limited data of this study, we suggest that one possible factor of rating severity could be the rater's multilingual proficiency, particularly in early childhood. It has been shown that multilinguals in Singapore develop skills to recognise various accents (Deterding and Poedjosoedarmo 2000). The rich exposure to various accents in Singapore might increase the processing speed of the aural input of accented speech (Weil 2001). If familiarity and exposure to a particular foreign accent results in a more lenient evaluation of the same accented speech (Winke and Gass 2013), it is possible that their ample exposure to various English accents might have influenced their L2 speech judgement. Again, further investigation is needed.

*Varieties of English*

The questionnaire data (Table 3, also see Appendix B) did not show notable differences among raters in the exposure to different varieties of English. The six raters reported two to five English varieties for each time point, including mainly Singapore English but also Philippine, Indonesia and other Asian English from their interlocutors such as their home helpers and friends, and American and British English through media from a very young age. Their exposure to different languages and English varieties continued throughout their school years. At school, they mingled with immigrant children or foreign students and teachers, which increased their exposure to different varieties of English.

The interview data also suggested that both Rater 3 and Rater 4 had ample exposure to multiple varieties of English at both home and school from a young age. Both raters commonly reported that they had a home helper at home who spoke Indonesian-accented English and both often watched American cartoon TV programmes. In addition, both the raters were exposed to other varieties of English at preschools. For example, Rater 3's class teacher (who used English) was from the Philippines. Although two raters both listed a number of English accents they were exposed to, both expressed difficulty in accurately quantifying it, particularly in their school period. Rater 3 stated:

We hear different accents all the time, like friends; we have students from China, Malaysia, other Asian countries. My close friend moved to Australia. She sometimes comes back and stay with me but she really sounds like Australian now . . . I spend many hours watching YouTube videos, mostly American sitcoms. But at school, some teachers are from England, so I'm fine with British English as well. So I don't even notice what accents I'm exposed to, most of the time.

This analysis suggests that Singaporeans are constantly exposed to various acoustic patterns in English (Tan 2012). It is possible that such complex experience of accented

English leads to some kind of perceptions towards particular accents, which might have some influence of judging foreign accents with similar phonetic patterns (Major et al. 2002). However, the data obtained in this study did not show any evidence for this. Further detailed investigation of the relationship between the multilingual raters' familiarity and perception towards different varieties of English and their judgement of the accented speech is needed.

**Conclusion**

The purpose of this study was to explore how raters judge L2 accentedness and comprehensibility and how their language background influences their judgement. The analysis showed the complex nature of language experience of English speakers in multilingual society. The study suggested some possible roles of language background in raters' evaluation of L2 speech as a matter for further investigation. First, the results suggested the tendency that raters who had better command of English over the other languages during their early childhood tended to judge the comprehensibility and accentedness of L2 speech in stricter terms. The findings indicate that a balanced exposure to two languages (English and Mandarin) at home in early childhood might have some impact on accentedness and comprehensibility judgement of L2 speech. This may be because, as Saito and Shintani (2016) suggested, raters who are more multilingual use a greater variety of resources including lexico-grammatical information as well as phonetic information to judge comprehensibility compared to those who are less multilingual. Previous research suggested that multilingual raters were more lenient than monolingual raters in their judgements (Saito and Shintani 2016) and that native raters' L2 backgrounds could influence their rating severity of non-native speech (Winke et al. 2013). The findings of the current study call for larger scale empirical studies to examine the correlations between rater severity and early childhood language exposure and proficiency.

We acknowledge that our study has several methodological limitations. First, the key constructs of this study – raters' exposure and proficiency – were assessed entirely relying on their retrospective report. Although a retrospective report is a common means for such purposes (e.g. Clyne, Rossi Hunt and Isaakidis 2004) and we believe that there was no other effective way to obtain the raters' record of language experience in the past, we could have tried to obtain data from multiple resources such as narrative frame or intensive interviews. Second, the proficiency scale in the questionnaire might not have been explicit enough for raters. The five points in the levels (beginner to native level) might not have been sufficiently self-explanatory to determine the raters' proficiency in different age ranges. The questionnaire would have been more valid and reliable if the proficiency scale had been presented with detailed description as to what they were able to do using the language. Also, an objective measurement of their proficiency in English and other languages would have made the results more robust (Thomas 1994). Finally, the questionnaire did not examine the raters' perceptions regarding accented speech. As discussed, to capture a holistic picture of an individual Singaporean's exposure to English varieties, it is necessary to investigate the raters' perceptions regarding, as well as exposure to, different accented English varieties (Winke et al. 2013). To overcome the above limitations and build further evidence, future research could investigate Singaporean children just before they enter primary schools. As the raters' different language exposures were mainly found in the pre-school period, investigating children immediately after this period would allow a more accurate picture of their language exposure and proficiency at that point. A larger-scale study would allow comparisons between groups of raters who had similar language exposures, as the findings of the current study indicated.

The strength of this study is that it takes the form of a case study, focusing on and examining in detail the judgement performances of six raters. Consequently, we could suggest the complex relationships between language experiences and native raters' judgement

of L2 accentedness and comprehensibility. As we observed from the interview data, multilingual individuals choose to use different languages depending on various factors. In real life, they may engage in code switching in different situations, depending on their language-mixing habits, their usual mode of interaction, the presence of monolinguals, the degree of formality, the form and content of the message uttered or listened to and the socio-economic status of communication partners (Grosjean 2004). Such a varied and changing linguistic profile certainly influences their language development, which might have some impact on their responses to accented speech. In order to capture such changing profiles, dynamic approaches to the study of multilingual language learning (De Bot, Lowie and Verspoor 2007) are needed.

**Notes**

1. A close analysis showed that Rater 3 was stricter in evaluating Task 3 than Task 2; in other words, Rater 3 judged speakers consistently.

**References**


Aman, N., P. Appleyard, A. James, M. Bte Roslan, T.K. Tan, and V. Vaish. 2007. *The linguistic ecology of Singapore students.* Presentation to teachers of SSS 2006 participating schools. Singapore: CRPP.

Barkaoui, K. 2013. "Multifaceted Rasch analysis for test evaluation." In *The companion to language assessment* (Vol. III: Evaluation, Methodology, and Interdisciplinary Themes, Part 10: Quantitative analysis, edited by A. Kunnan, 1301–1322, West Sussex, UK: John Wiley and Sons. doi:10.1002/9781118411360.wbcla070

Boersma, P., and D. Weenink. 2012. *Praat: doing phonetics by computer* [Computer program]. http://www.praat.org/.

Bradlow, A. and Bent, T. 2008. "Perceptual adaptation to non-native speech." *Cognition*, 106: 707–729.

Carey, Michael D., Robert H. Mannell, and Peter K. Dunn. 2011. "Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews?." *Language Testing* 28 (2): 201–219.

Cavallaro, F., and N. B. Chin. 2014. "Language in Singapore: From Multilingualism to English Plus." *Challenging the Monolingual Mindset* 156: 33.

Clyne, M., C. Rossi Hunt, and T. Isaakidis. 2004. "Learning a Community Language as a Third Language." *International Journal of Multilingualism* 1 (1): 33–52.

De Bot, K., W. Lowie, and M. Verspoor. 2007. "A Dynamic Systems Theory Approach to Second Language Acquisition." *Bilingualism: Language and Cognition* 10 (1): 7–21.

Derwing, T. M., and M. J. Munro. 2009. "Putting Accent in its Place: Rethinking Obstacles to Communication." *Language Teaching* 42 (4): 276–490. doi: http:0.1017/S026144480800551X

Derwing, T. M., and M. J. Munro. 1997. "Accent, Intelligibility, and Comprehensibility." *Studies in Second Language Acquisition* 19 (1): 1–16.

Deterding, D., and G. Poedjosoedarmo. 2000. "To What Extent Can the Ethnic Group of Young Singaporeans Be Identified from Their Speech?" In *The English Language in Singapore: Research on Pronunciation,* edited by A. Brown, D. Deterding, and E. L. Low, 1–9. Singapore: Singapore Association for Applied Linguistics.

Engelhard, G., and E. G. Stone. 1998. "Evaluating the quality of ratings obtained from standard-setting judges." *Educ. and Psychological Measurement* 58 (2), 179–196.

Gass, S. M., and E. M. Varonis. 1984. "The effect of familiarity on the comprehensibility of nonnative speech." *Language Learning*, 34 (1): 65–89.

Grosjean, F. 2004. "Studying Bilinguals: Methodological and Conceptual Issues." In *The Handbook of Bilingualism* edited by T. Bhatia and W. Ritchie, 32–63. Malden, MA: Blackwell.

Isaacs, T., and R. I. Thomson. 2013. "Rater Experience, Rating Scale Length, and Judgments of L2 Pronunciation: Revisiting Research Conventions." *Language Assessment Quarterly* 10 (2): 135–159. doi:10.1080/15434303.2013.769545

Jessner, U. 2008. "Teaching Third Languages: Findings, Trends and Challenges." *Language Teaching* 41 (1): 15–56.

Kennedy, S., and P. Trofimovich. 2008. "Intelligibility, Comprehensibility, and Accentedness of L2 Speech: The Role of Listener Experience and Semantic Context." *The Canadian Modern Language Review* 64 (3): 459–489.

Kim, Y-H. 2009. "An Investigation into Native and Non-Native Teachers' Judgments of Oral English Performance: A Mixed Methods Approach." *Language Testing* 26: 187–217. DOI: 10.1177/0265532208101010

Lalwani, A. K., L. May, and L. L. Kuah. 2005. "Consumer Responses to English Accent Variations in Advertising." *Journal of Global Marketing* 18 (3–4): 143–165.

Linacre, J. M. 2013. *Facets: Many-Facet Rasch-measurement* (Version 3.71.4) [Computer software]. Chicago: MESA Press.

Linacre, J. M. 2014. *Facets: Many-facet Rasch measurement* (Version 3.71.4) [Computer software]. Chicago: MESA Press.

Linacre, J. M., and B. D. Wright. 1994. "(Dichotomous mean-square) Chi-square fit statistics." *Rasch Measurement Transactions* 8 (2): 2360.

Lunz, M. E., and M. J. Linacre. 1998. Measurement designs using multifacet Rasch modeling. In *Modern methods for business research, edited by* G. A. Marcoulides, 47–77. Mahwah, NJ: Erlbaum.

Major, R. C., S. F. Fitzmaurice, F. Bunta, and C. Balasubramanian. 2002. "The Effects of Nonnative Accents on Listening Comprehension: Implications for ESL Assessment. *TESOL Quarterly,* 36 (2): 173–190. doi:10.2307/3588329

MacKay, I. R. A., J. E. Flege, and S. Imai. 2006. "Evaluating the Effects of Chronological Age and Sentence Duration on Degree of Perceived of Foreign Accent. *Applied Psycholinguistics* 27: 157–183. doi:10.1017/S0142716406060231

Ministry of Home Affairs. 2014. *2014 population in brief*. Singapore, The Ministry of Home Affairs. http://www.nptd.gov.sg/portals/0/news/population-in-brief-2014.pdf

Ministry of Manpower. 2015. *Number of Manpower*. Singapore Government. http://www.mom.gov.sg/documents-and-publications/foreign-workforce-numbers

Munro, M. J. (2008). "Foreign Accent and Speech Intelligibility." In *Phonology and Second Language Acquisition* edited by J.G. Hansen Edwards and M.L. Zampini, 193–218. Amsterdam: John Benjamins.

Munro, M. J., and T. M. Derwing. 2001. "Modeling Perceptions of the Accentedness and Comprehensibility of L2 Speech: The Role of Speaking Rate." *Studies in Second Language Acquisition, 23*(4): 451–468.

Munro, M. J., M. T. Derwing, and S. L. Morton. 2006. "The Mutual Intelligibility of L2 Speech." *Studies in Second Language Acquisition* 28 (1): 111–131.

Norris, D., McQueen, J. M., and Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology,* 47: 204–238.

Saito, K. 2011. "Beginner, intermediate and advanced Japanese learners of English in Canada." Unpublished corpus of second language speech.

Saito, K., and N. Shintani. 2016. "Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech?" *TESOL Quarterly,* 50 (2): 421–446. doi:10.1002/tesq.234.

Saito, K., P. Trofimovich., and T. Isaacs. 2015. "Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study." *Applied Linguistics*. doi:10.1093/applin/amv047

Tan, C. H., and A. F. Gupta. 1992. "Post-Vocalic /R/ in Singapore English." *York Papers in Linguistics* 16: 139–152.

Tan, P. K., and D. K. Tan. 2008. "Attitudes Towards Non-Standard English in Singapore." *World Englishes* 27 (3/4): 465–479.

Tan, Y. Y. 2012. "To R or Not to R: Social Correlates of/ɹ/in Singapore English." *International Journal of the Sociology of Language* 218: 1–24.

Thomas, M. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44: 307–336. doi:10.1111/j.1467-1770.1994.tb01104.x

Trofimovich, P., and W. Baker. 2006. "Learning Second Language Suprasegmentals: Effects of L2 Experience on Prosody and Fluency Characteristics of L2 Speech." *Studies in Second Language Acquisition, 28*(1): 1–30. doi:10.1017/S0272263106060013

Trofimovich, P., and T. Isaacs. 2012. "Disentangling Accent from Comprehensibility." *Bilingualism: Language and Cognition* 15 (4): 905–916. doi:10.1017/S1366728912000168

Weil, S. A. 2001. "Foreign-accented speech: Encoding and generalization." *Journal of the Acoustical Society of America.* 109: 2473 (A).

Winke, P., and 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. doi:10.1002/tesq.73

Winke, P., S. Gass, and C. Myford. 2013. "Raters' L2 Background as a Potential Source of Bias in Rating Oral Performance." *Language Testing* 30 (2): 231–252.

Yeoh, B., and L. Weiqiang. 2012. "Rapid Growth in Singapore's Immigrant Population Brings Policy Challenges." Migration Policy Institute. Retrieved from http://www.migrationpolicy.org/article/rapid-growth-singapores-immigrant-population-brings-policy-challenges

**Appendix A**
**Background Questionnaire**

| | |
|---|---|
| **Name** | |
| **Age** | |

**A. Family Languages**

| | | |
|---|---|---|
| Your parents' most dominant language | Father: | Mother: |

Please list all people who lived with you in your life (e.g. your family members and your helper). Indicate their English proficiency levels from:
**1. beginner, 2. intermediate, 3. high intermediate, 4. advanced, 5. native**

| Relationship with you (e.g. Father) | English proficiency (1-5) |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |

**B. The Use and Skills of Languages**

How often were you exposed to different languages at the three time points of your life: 5 (at the end of primary), 11 (at the end of primary) and 21 (now)? Please write the percentages for each language 1) at home, 2) at school (in the classroom), and 3) with friends, separately. In what occasions do you hear each language in your daily life? Please write down as many occasions as you can think of.

**Age 5**

| **Age 5** | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions (with whom? in what situations?)** |
| | **At home** | **At school** | **With friends** | |
| English | | | | |
| Mandarin | | | | |
| Other Chinese dialect ( ) | | | | |

| | | | | |
|---|---|---|---|---|
| Malay | | | | |
| Tamil | | | | |
| Other Indian dialect | | | | |
| Other language ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

Your language proficiency when you are 5 years old **(1. beginner, 2. intermediate, 3. high intermediate, 4. advanced, 5. native)**

| Lanugage | English | Mandarin | Tamil | Other ( ) | Other ( ) | Other ( ) | Other ( ) |
|---|---|---|---|---|---|---|---|
| Level | | | | | | | |

**Age 11**

| **Age 11** | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions (with whom? in what situations?)** |
| | **At home** | **At school** | **With friends** | |
| English | | | | |
| Mandarin | | | | |
| Other Chinese dialect ( ) | | | | |
| Malay | | | | |
| Tamil | | | | |
| Other Indian dialect | | | | |
| Other language ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

Your language proficiency when you are 11 years old **(1. beginner, 2. intermediate, 3. high intermediate, 4. advanced, 5. native)**

| Language | English | Mandarin | Tamil | Other ( ) | Other ( ) | Other ( ) | Other ( ) |
|---|---|---|---|---|---|---|---|
| Level | | | | | | | |

**Age 21 (Now)**

| **Age 21 (Now)** | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions (with whom? in what situations?)** |
| | **At home** | **At school** | **With friends** | |

| | | | | |
|---|---|---|---|---|
| English | | | | |
| Mandarin | | | | |
| Other Chinese dialect ( ) | | | | |
| Malay | | | | |
| Tamil | | | | |
| Other Indian dialect | | | | |
| Other language ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

Your language proficiency now **(1. beginner, 2. intermediate, 3. high intermediate, 4. advanced, 5. native)**

| Language | English | Mandarin | Tamil | Other ( ) | Other ( ) | Other ( ) | Other ( ) |
|---|---|---|---|---|---|---|---|
| Level | | | | | | | |

**C. Exposure to English Varieties (Accents)**

How often were you exposed to English with different accent at the three time points of your life: 5 (at the end of primary), 11 9at the end of primary) and 21 (now)? Please write the percentages for each language 1) at home, 2) at school (in the classroom) , and 3) with friends, separately. In what occasions do you hear different varieties of English in your daily life? Please write down as many occasions as you can think of.

| **Age 5** | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions** |
| | **At home** | **At school** | **With friends** | **(from whom? from what recourses? in what situations?)** |
| American English | | | | |
| British English | | | | |
| Standard Singapore English | | | | |
| Singlish | | | | |
| Other variety ( ) | | | | |
| Other variety ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

| Age 11 | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions (from whom? from what recourses? in what situations?)** |
| | **At home** | **At school** | **With friends** | |
| American English | | | | |
| British English | | | | |
| Standard Singapore English | | | | |
| Singlish | | | | |
| Other variety ( ) | | | | |
| Other variety ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

| Age 21 (Now) | | | | |
|---|---|---|---|---|
| **Language** | **Frequency (%)** | | | **Occasions (from whom? from what recourses? in what situations?)** |
| | **At home** | **At school** | **With friends** | |
| American English | | | | |
| British English | | | | |
| Standard Singapore English | | | | |
| Singlish | | | | |
| Other variety ( ) | | | | |
| Other variety ( ) | | | | |
| **Total** | **100%** | **100%** | **100%** | |

**Appendix B**
**Language profiles of the six participants**

| Age | Language background | R1 | R2 | **R3 ("Strict")** | **R4 ("Lenient")** | R5 | R6 |
|---|---|---|---|---|---|---|---|
| **5** | Family member's English proficiency (1-5) | Mother: 4<br>Father: 5<br>Helper: 5 | Mother: 4<br>Father: 5<br>Siblings: 5<br>Helper: 4 | Mother: 4<br>Father: 5<br>Siblings: 5<br>Helper: 4 | Mother: 2<br>Father: 5<br>Siblings: 5<br>Helper: 4<br>G-Mother: 0 | Mother: 3<br>Father: 2<br>Siblings: 5<br>Helper: 4 | Mother: 3<br>Father: 4<br>Siblings: 5<br>Helper: 5<br>G-Parents: 1 |
| | Number of language exposure | 2 | 3 | 3 | 4 | 3 | 3 |
| | Resource of language input | **English:** Father, Mother, siblings, helper, friends, TV<br>**Mandarin:** Mother (to teach Mandarin), between parents, friends, | **English:** Father, Mother, siblings, helper, friends, TV<br>**Mandarin:** Grandmother, mother's relatives, friends, Mandarin teacher<br>**Cantonese**: Grandfather, relatives | **English:** Father, Mother, siblings, helper, friends, TV<br>**Mandarin:** mother's relatives, Mandarin teacher<br>**Teochew:** Grandmother | **English:** Father, siblings, helper, friends, TV<br>**Mandarin:** Mother, siblings, friends, Mandarin teacher, TV, cousins<br>**Hokkein**: Grandparents, relatives | **English:** Mother, siblings, helper, friends, TV<br>**Mandarin:** Father, friends, cousins<br>**Tamil**: Friends | **English:** Father, siblings, helper, friends, TV<br>**Mandarin:** Mother, Father (when talked to Mother), friends, Mandarin teacher, TV<br>**Hokkein**: Grandparents |
| | Language exposure (%) | English: 85%<br>Mandarin: 15% | English: 60%<br>Mandarin: 35%<br>Cantonese: 5% | English: 90%<br>Mandarin: 8%<br>Teochew: 2% | English: 50%<br>Mandarin: 40%<br>Hokkein: 10% | English: 50%<br>Mandarin: 40%<br>Hokkein: 10% | English: 40%<br>Mandarin: 50%<br>Hokkein: 20% |
| | Own language proficiency (1-5) | English: 5<br>Mandarin: 2<br>Teochew: 1 | English: 5<br>Mandarin: 3<br>Cantonese: 1 | English: 5<br>Mandarin: 1 | English: 3<br>Mandarin: 4<br>Hokkein: 3 | English: 5<br>Mandarin: 4<br>Hokkein: 1 | English: 3<br>Mandarin: 4<br>Hokkein: 1 |
| | Exposure to English varieties (resources) | **Singapore** (family, friends teacher) | **Singapore** (family, friends teacher) | **Singapore** (family, friends teacher) | **Singapore** (family, friends teacher) | **Singapore** (family, friends teacher) | **Singapore** (family, friends teacher) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Indonesian** (helper) **American** (TV) | **Philippines** (helper) **American** (TV) | **Indonesian** (helper) **American** (TV) | **Philippines** (teacher) **Indonesian** (helper) **American** (TV) | **Philippines** (helper) **American** (TV) | **Philippines** (helper) **American** (TV) |
| **11** | Number of language exposure | 2 | 3 | 2 | 3 | 2 | 3 |
| | Resource of language input | **English:** Father, Mother, siblings, helper, friends, TV **Mandarin:** Mother (to teach Mandarin), between parents, friends, | **English:** Father, Mother, siblings, helper, teachers, friends, TV **Mandarin:** Grandmother, mother's relatives, friends, Mandarin teacher **Cantonese**: Grandfather, relatives | **English:** teachers, father, mother, siblings, helper, friends, TV **Mandarin:** friends, mother's relatives, Mandarin teacher | **English:** teachers, father, siblings, helper, friends, TV **Mandarin:** Mother, siblings, friends, Mandarin teacher, TV **Hokkein**: Grandparents, relatives | **English:** teachers, mother, siblings, helper, friends, TV **Mandarin:** Father, friends | **English:** father, siblings, helper, friends, teachers, TV **Mandarin:** Mother, Father (when talked to Mother), friends, Mandarin teacher, TV **Hokkein**: Grandmother |
| | % of language exposure | English: 90 Mandarin: 10 | English: 70 Mandarin: 20 Hokkein: 10 | English: 80 Mandarin: 20 | English: 70 Mandarin: 20 Hokkein: 5 Tamil: 5 | English: 70 Mandarin: 30 | English: 80 Mandarin: 15 Hokkein: 5 |
| | language proficiency | English: 5 Mandarin: 2 | English: 5 Mandarin: 4 | English: 5 Mandarin: 3 | English: 5 Mandarin: 5 Hokkein: 3 | English: 5 Mandarin: 4 | English: 5 Mandarin: 5 Hokkein: 1 |
| | Exposure to English varieties | **Singapore** (friends, teacher, family, relatives) **Indonesian** (helper) | **Singapore** (friends, teacher, family, relatives) **Philippines** (helper) | **Singapore** (friends, teacher, family, relatives) **Philippines** (helper) | **Singapore** (friends, teacher, family) **Philippines** (helper) | **Singapore** (friends, teacher, family, relatives) **Philippines** (helper) | **Singapore** (friends, teacher, family, relatives) **Philippines** (helper) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **American** (TV) | **American** (TV) | **American** (TV) | **American** (TV) | **American** (TV)<br>**Chinese** (friends) | **American** (TV) |
| **21** | # of language exposure | 2 | 2 | 2 | 3 | 2 | 2 |
| | Resource of language input | **English:** friends, teachers, parents, YouTube, TV<br>**Mandarin:** some friends | **English:** friends, teachers, parents, YouTube, TV<br>**Mandarin:** some friends | **English:** friends, teachers, roommates, YouTube, TV<br>**Mandarin:** some friends | **English:** friends, teachers, roommates, YouTube<br>**Mandarin:** Mother, friends<br>**Hokkein**: Grandparents, relatives | **English:** friends, teachers, YouTube<br>**Mandarin:** friends, roommates (from China) | **English:** friends, teachers, roommates, YouTube<br>**Mandarin:** parents, friends, TV with family |
| | % of language exposure | English: 90<br>Mandarin: 10 | English: 80<br>Mandarin: 20 | English: 90<br>Mandarin: 10 | English: 80<br>Mandarin: 10<br>Hokkein: 5<br>Tamil: 5 | English: 70<br>Mandarin: 30 | English: 80<br>Mandarin: 20 |
| | Language proficiency | English: 5<br>Mandarin: 3 | English: 5<br>Mandarin: 4 | English: 5<br>Mandarin: 3 | English: 5<br>Mandarin: 5<br>Hokkein: 1 | English: 5<br>Mandarin: 4 | English: 5<br>Mandarin: 5<br>Hokkein: 1 |
| | Exposure to English varieties (resources) | **Singapore** (friends, teacher, family)<br>**Chinese** (friends)<br>**Australian** (friends)<br>**American** (online videos)<br>**Malay** (friends) | **Singapore** (friends, teacher, family)<br>**Chinese** (friends)<br>**Malay** (friends)<br>**American** (online videos, music)<br>**British** (online videos, aunt) | **Singapore** (friends, teacher, family)<br>**Chinese** (friends)<br>**Philippines** (friends)<br>**American** (online videos)<br>**British** (teacher) | **Singapore** (friends, teacher, family)<br>**Chinese** (friends)<br>**American** (online videos)<br>**British** (online videos) | **Singapore** (friends, teacher, family)<br>**Chinese** (roommate)<br>**German** (friend overseas)<br>**American** (online videos, music)<br>**Indian** (friends) | **Singapore** (friends, teacher, family)<br>**Malay** (friends)<br>**Chinese** (friends)<br>**American** (online videos, teacher)<br>**British** (teacher) |