

Accent and academic listening assessment: A study of test-taker perceptions

Luke Harding

The University of Melbourne

Abstract

This study investigated test-taker perceptions of an academic listening test with diverse accented speakers. A listening test was constructed with speakers representing four first-language backgrounds: Australian English, Mandarin Chinese, Japanese and Bengali. This test was administered to 44 ESL learners, together with a questionnaire designed to elicit participants' accent preferences. Following this, test-taker perceptions were explored through four focus groups with learners at different proficiency levels, as well as three one-on-one interviews. Test-taker perceptions were elicited on three key issues: accent-related difficulty, a shared-L1 advantage in comprehension, and test fairness. Results showed that lower proficiency listeners perceived accent-related difficulty in different ways to higher proficiency learners; that a shared-L1 appeared to distract listeners rather than aid their comprehension of a speaker; and that views about the fairness of an academic listening test with diverse accented speakers were complex and varied across the cohort. Implications of these findings for future research are discussed.

1 Introduction

Owing to the spread and diversification of English in outer- and expanding-circle contexts, and the emergence of English as a Lingua Franca, debate has surfaced in language testing literature over whose “norms” should take primacy in English language assessment standards (Davies, et al, 2003; Lowenberg, 1992; 1993; 2002). Davies, Hamp-Lyons and Kemp (2003) have characterised opposing sides of the debate broadly as those who support the testing of International English (IE) – a monolithic, globalised standard (see Quirk, 1990) – and those who hold a World Englishes view (see Kachru, 1986), and support the recognition and incorporation of localised varieties of English. This debate has spawned research and comment on the validity of testing particular cohorts against standardised, primarily “inner-circle” norms (Brown & Lumley, 1998, Hamp-Lyons & Davies, 2006; Lowenberg, 1992; 1993), however this literature has tended to focus on the assessment of speaking and writing. Less has been said of the particular challenges for the assessment of listening where the representation of accent varieties in recorded input has also come into question.

Large-scale tests of academic English such as TOEFL and IELTS, in their most recent incarnations, appear to acknowledge the changing communicative needs of candidates through their approach to speaker accent in listening assessment. In the TOEFL iBT, for example, innovations to the listening section have included the representation of a range of global accents. Yet this range extends from North American to British and Australian accents, and these accents will only be heard on one part of the listening section (mini-lectures), and may or may not appear on any given listening test (see ETS, 2005). Cambridge ESOL makes similar claims regarding their approach to international communication; the *IELTS Handbook* states that the listening paper utilises “a range of English accents and dialects ... in the recordings which reflects the international usage of IELTS” (2006, p.6). In practice

this range of accents appears to be limited to British, Australian, North American or New Zealand varieties (see Cambridge ESOL, 2008).

While conservative, these changes have largely been designed to increase authenticity in the listening component by more accurately representing the target-language use (TLU) domain of academic institutions. According to Taylor (2006), the guiding principle behind the IELTS approach has been, "that the test should represent those English varieties used in the contexts in which IELTS test-takers are likely to find themselves, i.e. the 'dominant host languages'" (p.56). However, students at an academic institution in a modern, metropolitan setting are likely to encounter lecturers, tutors or teaching assistants with a wide range of accents including non-native varieties. To achieve greater authenticity, academic listening tests should ideally also include speakers with such accents in oral stimuli for listening assessment. Furthermore, the inclusion of greater accent diversity in assessment practices may create positive washback. Consider the effect that using exclusively inner-circle, native-speaker accents on major listening tests is having on policy decisions in the wider ESL/EFL community such as the selection of speakers for course-books and low-stakes tests within language programs, the attitudes of learners towards teachers with non-standard accents and, at the extreme, the hiring of NNS language teachers (see Lippi-Green, 1997).

Yet while the inclusion of a broad range of accents in EAP listening assessment may be viewed as beneficial, it also raises a number of problematic issues related to fairness and test usefulness. Firstly, in the use of diverse accented speakers there exists the potential for severe breakdowns in comprehension, or for listeners to find particular accents unreasonably difficult to understand. This concern leads Buck (2001, p.35) to caution that in a listening test, "accent is a very important variable ... [that] can cause problems and may disrupt the whole listening comprehension process." Secondly, there is potential for test bias where listeners who share a speaker's L1 are advantaged over others when listening to that speaker. This concern was addressed in the context of listening assessment by Major et al. (2002) who warn that a bias effect

may exist, though their findings are inconclusive. Thirdly, there is potential for test-takers to perceive a listening test with diverse-accented speakers as unfair. Taylor (2006) suggests that although many applied linguists have a progressive stance in the debate over appropriate norms and models, “we should not ignore or override the attitudes and perceptions of learners themselves” (p.52).

2 Research objectives

Owing to the scarcity of research on accent and listening comprehension from the specific perspective of language assessment, and in particular the lack of any qualitative evidence of test-taker perceptions, this study aimed to explore the issues outlined above through administering a short EAP listening test with diverse-accented speakers to a cohort of ESL learners. Three specific research questions were posed:

- 1) What are test-takers’ perceptions of accent-related difficulty?
- 2) What are test-takers’ perceptions of a shared-L1 advantage?
- 3) What are test-takers’ perceptions of the fairness of an academic listening test with diverse-accented speakers?

3 Methods

3.1 Speakers

The University of Melbourne was canvassed for potential speakers with a wide range of accents. The only pre-requisite for speakers was that they had attained or were studying for a graduate degree. This was so that speakers would accurately represent authentic lecturers or tutors in an academic institution. Nine initial speakers were identified; their profiles are presented in Table 1.

Table 1 - Speaker profiles

	Gender	Age	Nationality	L1	LoR	Accent self-report
A	F	28	Singaporean	English/Mandarin Chinese	10 months	Singaporean
B	F	34	Australian	English	34 years	Australian
C	M	36	Australian	English	36 years	Australian
D	F	31	Japanese	Japanese	14 months	Japanese (slightly American)
E	F	28	Chinese	Mandarin Chinese	3 months	American/Australian – basically Chinese
F	F	24	Chinese	Mandarin Chinese	4 months	Chinese
G	F	30	Japanese	Japanese	2.5 months	Japanese British English
H	M	40	Tanzanian	Swahili	6 months	African
I	M	37	Bangladesh	Bengali	5 months	American/British

[L1 = first language; LoR = length of residence in Australia]

Nine scripts on general academic topics were selected from the *TOEFL® Test Preparation Kit* (ETS, 2000) and speakers were asked to record one script each. All scripts were monologic, and were between 200 and 300 words in length. The speakers were encouraged to practice the script if they wished, but the focus of recording was on eliciting an authentic speaking performance of the scripted material. Occasionally, speakers made mistakes and re-read selected paragraphs or sentences. The recordings were made using an mp3 iRiver recorder with an external microphone, and all recordings were then uploaded to the Sound Studio 2.0.7 digital audio editing software program, which enabled post-hoc editing. The sound files were copied to a CD, which comprised the nine edited oral stimuli separated by pauses.

In order to select four speakers of different accents from the initial pool of nine, two raters provided evaluations of each speaker. Both raters were native speakers of English (British and Australian), and both were language professionals. The raters listened to all nine speakers and

gave impressionistic judgements of each speaker with the aim of selecting one Australian English speaker, one Mandarin Chinese accented speaker, one Japanese accented speaker, and one speaker from the two remaining varieties. Judgements were based on which speakers sounded like authentic lecturers and which were reasonably comprehensible. This process was by no means exhaustive, and its limitations are discussed later in the paper. The selected speakers are presented in Table 2:

Table 2 - Selected speakers

Speaker #	Speaker accent	Code from Table 1	Gender	Nationality
1	Australian English	B	Female	Australian
2	Mandarin Chinese	F	Female	Chinese
3	Japanese	G	Female	Japanese
4	Bengali	I	Male	Bangladeshi

3.2 Test construction

A “World English accents” listening test was constructed from the four scripts previously read by the four speakers identified in Table 2. In the *TOEFL® Test Preparation Kit* (ETS, 2000), each script is accompanied by four multiple-choice items, and these were used as the test items. Thus, the test consisted of four “mini-tasks” (scripts plus accompanying items), and 16 items in total. To overcome the potential problem that each of the mini-tasks may be of unequal difficulty, information provided in the *TOEFL® Test Preparation Kit* (ETS, 2000) was used to select sets of items that were of approximately equal difficulty. It was found that average item difficulty for each of the four mini-tasks chosen was between 0.55 and 0.57. The topics of the particular scripts chosen were: a lecture on

early insurance policies, a description of a space-suit, a lecture on the Earth's interior, and a lecture on the history of mail delivery in the United States.

The test was designed to be delivered in the TOEFL Paper-Based Test format. In this format, as listeners hear oral stimuli, they are not allowed to write or take notes anywhere on the test paper. Following each stimulus, listeners are given question prompts, and must select their answer from multiple-choice tasks presented in an answer booklet. Responses are recorded on a separate answer sheet.

3.3 Listeners

Forty-one adult ESL students were drawn from an ELICOS (English Language Intensive Courses for Overseas Students) program at a language centre in Melbourne. In addition, three postgraduate university students were included in the sample. The proficiency level of the ESL students ranged from pre-intermediate to upper-intermediate/IELTS preparation. The university students were of a higher level still, and could be classified as advanced. 72% of the participants were female and 28% male; the average age was 26.5. The listeners came from a range of L1 backgrounds, though a large majority were Spanish speaking Colombians; the representations of first language groups are provided in Table 3:

Table 3 - Listeners by first language

L1	Frequency (n = 44)
Spanish (Colombian)	19
Korean	10
Japanese	8
Chinese	3
Thai	2
Vietnamese	1
Sinhalese	1

3.4 Procedure

The test was administered in four classes at the language centre. It was also administered individually to the three university students at the University of Melbourne. In all cases, the testing procedure was identical: listeners completed the test and were then asked to fill in a post-test questionnaire which collected information on their language and educational background, as well as their attitudes towards accents in the classroom and in their own speech. This questionnaire was given after testing so as not to flag “speaker accent” in the minds of the participants as they listened.

Focus groups were held with participants following the administration of each test at the language centre, and individual interviews with each of the three university students.¹ A brief set of guide questions for the focus groups and interviews was developed to

¹ Although the focus groups were larger than the optimal number at 7-12 people, it was unavoidable as testing had taken place during regular class time.

elicit perceptions from participants on the three issues relevant to the research questions: whether, and to what extent, the accents of speakers contributed to difficulty; whether, and to what extent, a shared-L1 was an advantage; and whether candidates perceived a listening test with different accents to be fair. Other related issues which emerged during the focus groups were also pursued.

4 Analysis of test scores and questionnaire responses

Following the testing procedure, tests were scored and these data, along with answers on the post-test questionnaire, were entered into SPSS. Some basic statistical analyses were carried out on these data sets to provide context for the analysis of perceptions from the focus groups and interviews. Firstly, the means and standard deviations were calculated for whole group performance with each of the four speakers, and these are shown in Table 4:

Table 4 - Descriptive statistics for whole group performance with accented speakers

Speaker accent	N	Mean ²	Standard deviation
Australian English	44	1.48	1.05
Mandarin Chinese	44	1.52	1.24
Japanese	44	1.68	1.05
Bengali	44	1.25	0.94

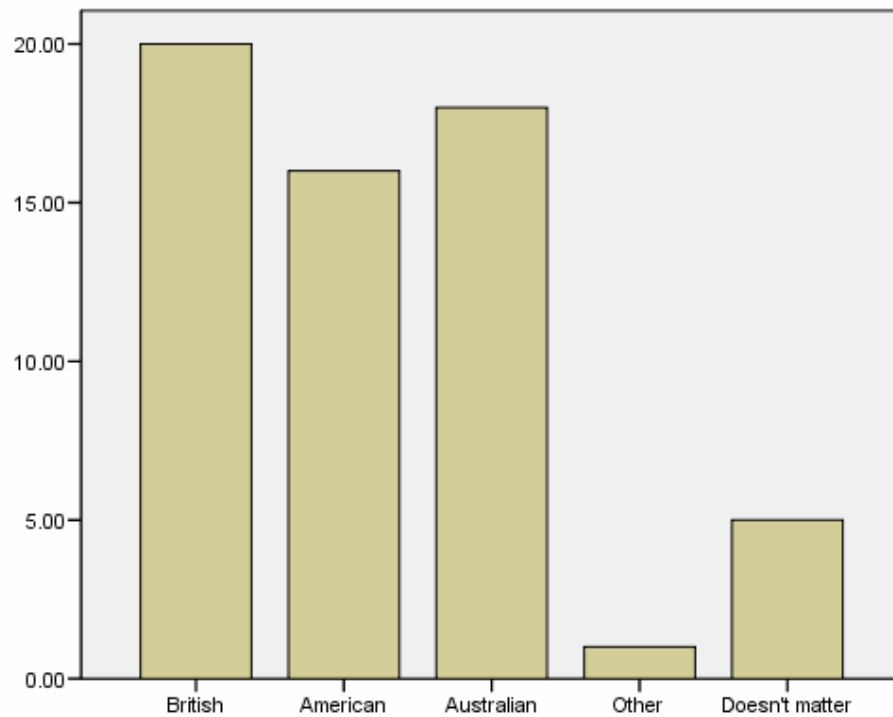
A repeated measures ANOVA suggested that there were no significant differences in the whole group's performance with different accented speakers [Wilks' Lambda = 0.89, $F(3, 41) = 1.68$, $p > 0.05$]. However this

² There was a total available score of 4 on each of the tests.

analysis should be treated with caution because, although each sub-test in the research design contained items of roughly the same average difficulty (see section 3.2), the tests were still not demonstrated to be of a suitably equivalent difficulty to isolate speaker-accent as the only within-subjects factor. What we might infer from the descriptive statistics, however, is that there was not a fundamental breakdown in comprehension across the whole group with any of the speakers.

Participants were also asked on the post-test questionnaire to nominate the accents they prefer their teachers speak, and the accents they are aiming for when they speak English. Figure 1 shows that, for preferred teacher accents, there was a strong liking for inner-circle, native speaker varieties. British and Australian accents were most preferred, though some participants indicated that it did not matter what variety a teacher had. One participant, who was Japanese, nominated a Japanese accent as her preferred accent in which to be taught. Follow-up comments suggested that the reasoning behind choices was varied: British accents tended to be preferred on intelligibility and aesthetic grounds (e.g., "it's easy to understand and I like it"; "just I like the way they talk"), whereas the preference for Australian English was often based on pragmatic grounds (e.g., "because I live in Australia and I want to study in Australia").

Figure 1 – Preferred teacher accents

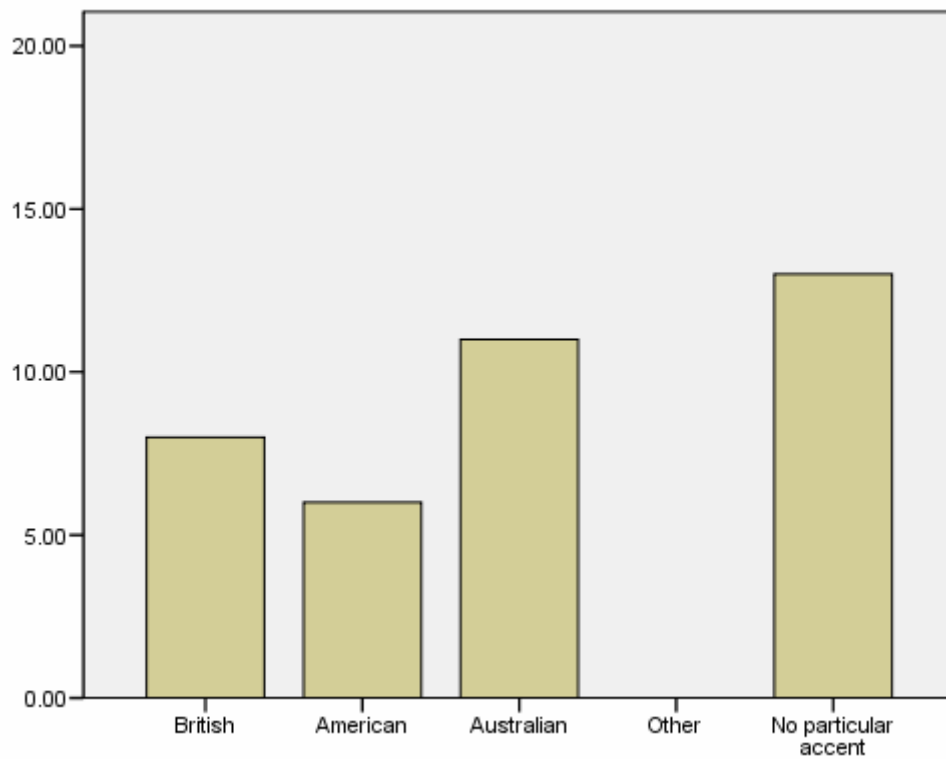


The response to the question concerning personal accent “aim” followed similar trends, though, as Figure 2 illustrates, with a good number of participants indicating that they were aiming for no particular accent. Again, the reasoning behind the choice of native speaker accent targets was a mix of aesthetic preference and the practicality of learning the accent of the host community. However, several participants seemed to hold more proficiency-oriented goals for their speaking (e.g. “just I want to speak English fluently”).

This questionnaire data suggests that the cohort of participants exhibited a general preference for standard, native-speaker accents for

teachers, but were also mindful of a distinction between the acquisition of accent and the ability to communicate effectively. These underlying attitudes are useful in establishing a context for interpreting the findings of the focus groups and interviews that are presented in the following section.

Figure 2 – Own accent goals



5 Focus groups and interviews

Recordings of the focus groups and the individual interviews were transcribed and analysed for content under three themes: perceptions of difficulty, perceptions of own-accent advantage, and perceptions of fairness.

5.1 Speaker accent as a source of difficulty

Speaker accent was mentioned across all focus groups as a source of difficulty; however there were some marked differences in perceptions of the impact of accent at higher and lower proficiency levels. Excerpt 1, from the pre-intermediate class focus group, demonstrates that for students at this level, speech rate and academic vocabulary are equally salient features of difficulty as accented speech:

Excerpt 1 (Pre-intermediate class focus group)

- I: My first question is how was the test?
SS: Difficult (many voices) (laughter)
I: Okay ... okay why was it difficult?
S1: Because it was fast ...
I: Fast?
S2: Difficult words ...
I: Words ... okay
S3: The accent ... the first accent [Australian English] was more difficult than the other, in the other interviews

By contrast, the highest-level class – the IELTS preparation class – immediately identified speaker accent as the “difficult element” of this listening test:

Excerpt 2 (IELTS preparation class focus group)

- I: Some people said you found it difficult .. what was difficult about it?
S2: The accents
S3: There were many accents
S4: Different kinds of accents, yeah ..
I: Different accents?
S2: Yeah
I: Did everyone find the different accents difficult?
SS: Yes (many voices)

That speaker accent is not as salient for lower proficiency listeners is supported to some extent by Goh (1999) who found that low and high proficiency learners reported different factors that they perceived affected their listening ability, with the high ability group mentioning speaker accent and speaker competence amongst twelve other factors affecting their comprehension, while the low ability group reported only four factors: vocabulary, speech rate, type of input (text) and prior knowledge (Goh, 1999, p. 34).

When the lower proficiency group did mention accent, their statements reflected a different perception of which accent was most difficult compared with the higher-level groups. Excerpt 3 shows that some members of the pre-intermediate group found the first accent (Australian English) more difficult than the fourth accent (Bengali). This low proficiency group was the only group to report this:

Excerpt 3 (Pre-intermediate class focus group)

- S3: The accent ... the first accent [Australian English] was more difficult than the other, in the other interviews
I: The first one was more difficult?
S3: Yes, for me ...
S4: For me it's the second one [Mandarin Chinese]...

- SS: Yes the second one (several voices)
 S5: The last one [Bengali]...
 I: You think the last one ... okay
 SS: The last one was the best ...
 I: The last one was the best?
 SS: Yes (several voices) (laughter)

These impressions might be explained by the difference in speech rate between the four speakers. A fairly crude measure of words-per-second shows that, of the four speakers, the Australian English speaker was the fastest (Table 5):

Table 5 - Speech rate (words/second)

Australian English	Mandarin Chinese	Japanese	Begali
2.98	1.99	2.23	2.23

However, this measure also shows that the Mandarin Chinese speaker was the slowest – and several pre-intermediate students reported that they also found that speaker difficult to understand. It is also possible that the impression of difficulty for the lower level test-takers with the Australian English accented speaker was due to the fact that English is a stress-timed language, whereas the bulk of participants from the pre-intermediate class were L1 speakers of syllable- or mora-timed languages (Spanish, Japanese, Korean). This may have added to their impression that the Australian English speaker was speaking quickly.

In the higher level classes, a distinction was generally made between the first accent (Australian English) and the others, with the native-speaker accent perceived as clear, and a range of views surrounding which of the remaining accents was most difficult. For example, in the intermediate class, the Australian English speaker was singled out as more comprehensible, and the Bengali speaker was generally perceived as the most difficult to understand:

Excerpt 4 (Intermediate class focus group)

- I: Okay, first question, how did you find that test?
S1: Ah, in my personal opinion, I found it a little bit hard ...
I: Okay
S1: Because, it's, ah, a mix, ah, of different accent ...
I: Okay
S1: So maybe the first one [Australian English] was .. very ah clear for me
I: Uh huh
S1: The second one [Mandarin Chinese] a little bit hard(er)
I: Right
S1: And the last one [Bengali] it was hard, the hard one ...
SS: Yes (many voices)
S2: Yes, the last one was very hard ...
I: The last one was the most difficult?
SS: Yes (many voices)

At first glance the upper-intermediate class appeared to find all speakers difficult (see excerpt 5). However the first speaker is again separated from the other three speakers with the phrase "but the other three maybe", and the difficult accent(s) is identified as "Indian":

Excerpt 5 (Upper-intermediate class focus group)

- I: Did you find um any of the speakers difficult to understand?
S3: All of them ...
S4: Yes ... because I know it's not local language ...
SS: Yeah, yeah (many voices)
S5: First people maybe she's Aussie or ... England ...
S6: American ...
S5: But the other three people maybe

S2: Indian!
SS: (laughter)

While the IELTS preparation class also single out the Australian English speaker as easier to understand, they appear to find the Mandarin Chinese and Japanese accented speakers more difficult:

Excerpt 6 (IELTS preparation class focus group)

I: Some people said you found it difficult .. what was difficult about it?
S2: The accents
S3: There were many accents
S4: Different kinds of accents, yeah ..
I: Different accents?
S2: Yeah
I: Did everyone find the different accents difficult?
SS: Yes (many voices)
S2: The second one [Mandarin Chinese]
S4: Absolutely
I: The second one?
S4: The second, yes
S5: The third one [Japanese], I think
I: The third one
S4: No, the third ...
S6: I think the first one [Australian English] is the most ... the easiest one
S4: Easiest
S5: Yeah, it's th(e) easy
I: Okay, do you know what the first accent was?
S1: British, I think

The higher proficiency students, who had previously nominated speaker accent as a major source of difficulty in the test, were also able to supply a range of reasons for the perceived accent “cost”. For example, one student in the upper-intermediate class pointed specifically to the pronunciation of the Mandarin Chinese-, Japanese- and Bengali-accented speakers as the sources of comprehension difficulty:

Excerpt 7 (Upper-intermediate class focus group)

S1: The last three people I really uh I can understand nothing because they speak ... I think, some words they don't pronounce all the words correct, so I seem, I can confuse, oh: .. I guess ..

Michiko – an advanced-level Japanese university student – alluded to the extra work required to process words and sentences while listening to some of the accented speakers on the test:

Excerpt 8 (Michiko: Japanese university student)

M: Yeah because of the accent, or because of the pronunciation I spent I had to spend more time

I: Mm

M: Trying to figure out what what this word was instead of what this sentence was

These comments suggest that accent was perceived – at least by higher proficiency test-takers – as having an effect on word recognition, and on processing time.

5.2 Perceptions of a shared-L1 advantage

Although the numbers of Japanese and Mandarin Chinese listeners in the sample precluded any statistical measures of own-accent advantage, the focus groups revealed that, at least as perceived by listeners, there was no particular advantage in listening to their own-accent. Excerpt 9 demonstrates that, firstly, an own-accent advantage might be dependent to some extent on first recognising an accent as one's own:

Excerpt 9 (Intermediate class focus group)

- I: Is anyone here a Mandarin Chinese speaker?
SS: [S7] is Chinese ...
I: So did you find the second speaker easy to understand?
S7: Uh: .. I don't think the second one .. I, I don't know the second one is from China (laughs)

Even when a shared accent was recognized, it was not necessarily viewed as more comprehensible. Excerpt 10 shows that other speech characteristics beyond a familiar accent are important in perceptions of intelligibility:

Excerpt 10 (Upper-intermediate class focus group)

- I: Did you find the second speaker easy to understand [to Mandarin Chinese speaker]
S8: Uh: ... no ...
SS: (laughter)
I: No?
S8: Er: ... because she speak first sentence, I can know she is uh Chinese .. so maybe pronouncing is the same but, uh: ... I think maybe uh she didn't, didn't uh: practice so uh always stop ..

S8's comments signal the importance of speaker qualities, such as fluency, to the perception of speaker comprehensibility – a variable that was not controlled adequately at the speaker selection stage. The listener also expresses a negative attitude towards the speaker's language ability, and positions the speaker as a "learner" who needed to do more practice, indicating that she did not "believe" the speaker's assumed role as a lecturer.

This tendency to critically evaluate a speaker from a shared-L1 perspective was also a feature of the Japanese university level students' comments on the Japanese speaker. In Excerpt 11, Michiko appears to have been distracted by the speaker's approximations of /r/ and /l/:

Excerpt 11 (Michiko: Japanese university student)

- I: And how about the Japanese accent, did you feel that one was easier to listen to?
- M: It was more difficult for me actually ...
- I: Really?
- M: Yeah, I thought ... I um ... I thought she had some kind of American accent, rather than an Australian accent
- I: Okay
- M: Yeah
- I: And that sort of affected your understanding ...
- M: Lots of /r/ sounds and ...
- I: Okay, 'cause the Australian accent obviously we don't say final /r/
- M: Mm, and because she confused between /l/ and /r/ sound as well
- I: Mm
- M: So I wasn't .. sometimes I wasn't sure what she was trying to say ... I could .. I probably could guess, but if I spent time on guessing what she was saying I then I would lose what she was talking about
- I: Mm
- M: You know what I mean?

In a separate interview, Kazumi, another Japanese university student echoes Michiko's remarks, this time explicitly stating that she was distracted by the pronunciation of these particular sounds:

Excerpt 12 (Kazumi: Japanese university student)

- K: And ... yeah the third speaker was, I think she's Japanese
I: Right .. yeah
K: And
I: Yes, you're correct actually
K: She: and she has Japanese accent and American accent
I: Uh huh
K: Quite strong
I: Uh huh
K: But I guess ... that I still got distracted by her like especially /r/
sound and it's I think it's common for Japanese speakers who,
like who are very conscious about pronunciation they tend to
focus on /r/ and /l/ sound
I: Mm mm
K: And because they're so conscious about /r/ and /l/ they tend to ..
overdo it

It is clear from the participants' comments that the relationship between a listener and a speaker who share an L1 is potentially complex. The Japanese and Chinese L1 participants from the low proficiency groups seemed unable to identify particular speaker accents, indicating that whatever advantage they may have gained with these speakers was not noticed at a conscious level. The comments from the university level learners show that noticing a shared accent may in fact be more of a distraction than an advantage, which reflects the findings of Fayer and Krasinski (1987) that NNSs are typically harsher in their judgment of common-L1 speakers than are others.

5.3 Perceptions of fairness

The final topic of interest concerned the fairness of a listening test with diverse speaker accents. At all proficiency levels, focus group and interview data yielded a wide range of views on the general fairness of such a test. These assessments were made, however, with reference to different aspects of what constitutes “fairness”. For example, in Excerpt 13, members of the Intermediate class discuss the test’s fairness in relation to its increased difficulty, and its relevance to the needs of a language learner:

Excerpt 13 (Intermediate class focus group)

- I: Do you think that it’s fair to use different accents on a listening test?
S2: In IELTS?
I: In any test, in any academic listening test ...
S2: Oh my God ... it’s too difficult no
S1: No ...
I: You said not fair?
S1: I think it would be better in my opinion to use just the main accent like the Australian accent ...
S3: Yeah
S1: ... because we are ... doing ... we are doing English .. Australian English .. so we need to learn Australian English and we need to be tested in Australian English

In this example, the “fairness” of a test with diverse accented speakers is evaluated through the lens of “fitness for purpose”³; the notion that the test should match the context in which the participants find themselves.

³ This phrase is used by Taylor (2006) in describing the importance of contextual relevance to test validity.

Similarly, “fairness” was evaluated in light of whether this difficulty that test-takers experienced with accent was of relevance to the construct under test. For example, the test-taker in Excerpt 14 appears to regard the ability to cope with accent varieties as construct-irrelevant in his current, ESL context:

Excerpt 14 (IELTS preparation class focus group)

S13: Uh: I think it's not fair because as ah, as he told you when you go to a place where they speak English you, maybe you are expecting to listen just native speakers .. an: when you hear [non-native speakers] in a test I think they don't really evaluate or test your real life knowledge in English because you need longer to get accustomed to it ...

However a number of other test-takers saw a direct connection between the presence of diverse accents on the test, and their “real-life” language needs, as illustrated in Excerpts 15, 16, 17 and 18:

Excerpt 15 (Pre-intermediate class focus group)

I: You think it's fair?
S1: Yeah, because ... you are going to use English to speak, not only with English speakers ...
S2: Yes, that's true ...
S1: English speakers ... everywhere

Excerpt 16 (IELTS preparation class focus group)

I: Do you think that it's fair to use different accents on a listening test, like this .. I'll start with you, what do you think?

S1: Oh yes yes yes yes .. the reason is because you need, you need listening in all kinds of accents because the real life involves all those kinds of accents ..

Excerpt 17 (IELTS preparation class focus group)

S3: Yes well English is spoken for different countries, different cultures so you need to be accustomed to different accents in works .. in different kinds of works, in jobs and studies in general

Excerpt 18 (Kazumi: Japanese university student)

K: Um: ... I'm not sure about fairness ... um ... I mean in real life you will deal with many accents like different variety of English so in a way it's fair because real thing

I: Mm

K: Like you wouldn't just, even you come to Australia you wouldn't just talk to Australian English speakers you would have to deal with people from like different like um linguist(ic) linguistic backgrounds

One test-taker expressed the opinion that the ability to cope with various accents *was* the measure of his listening ability:

Excerpt 19 (Upper-intermediate class focus group)

S6: ... I thought that, that all these accents are good, because, because you have to understand all them, it's very important because if you can understand Japanese and Indian or a different nationality .. it's the only way that you can know that you: have improved your language ..

Thus when fairness was framed as “fitness for purpose” or “construct relevance”, the data shows that test-takers did not express a uniform view. Initially, this seems at odds with the preliminary data collected through the questionnaire which indicated that participants overwhelmingly preferred that their teachers have British, Australian or American accents. However it is clear that, for this cohort, there is a distinction made between those accents which are seen as desirable in a classroom context – the most intelligible or aesthetically attractive – and those which are perceived as necessary in a testing context – meeting the “real-life” communicative needs of the candidate.

Fairness was also framed in terms of “access” to accent: a test-taker’s familiarity with certain accents, and their likelihood of having encountered other accents before:

Excerpt 21 (IELTS preparation class focus group)

S11: Yeah I think that could be unfair if you are taking a test because there were, there are many many accents so in a test you can’t catch all of them, and also your teacher usually are from Australia in this case, so you used to, to listen in Australian accents

In a similar vein, two comments related to the potential for an own-accent advantage:

Excerpt 22 (IELTS preparation class focus group)

S7: I think it’s unfair because ah it is easy for, for me to understand the Japanese accent

I: Okay

S7: So I think it’s unfair ...

Excerpt 23 (Upper-intermediate class focus group)

S3: I think depends on their nationality .. maybe I live in Korea, and speaker's Korean .. but I'm benefit of that, but the other person is not, so ... it's not fair I think

Finally, accented speech was construed as an anxiety-inducing element in the particular context of an examination. One participant expressed the view that creating such anxiety would be unfair:

Excerpt 24 (IELTS preparation class focus group)

S4: I think that's unfair because you have to add some, another, some other factors like you are eh stress and the condition of exam, and I think if erm you feel confide(ent), you don't feel confident and so you know that sort of accent are completely erm different for you .. I think that could be eh: a terrible experience .. I don't think so ...

Similarly, Kazumi noted that anxiety may be heightened when accents on a listening test do not match the expectations of candidates:

Excerpt 25 (Kazumi: Japanese university student)

K: I I suppose you have to deal with different kinds of English so
I: Mm
K: It's ... it's a good thing to have tests with different accents
I: Mm
K: But if you give this test to Japanese
I: Mm
K: Student in Japan they will be like panicking and they will be like "oh I don't understand" it's just, it's only because of the accents,

because I'm not familiar with those accents like they will say "oh we're educated in American English so it's not fair"

I: Yep

K: "We're not familiar with those Englishes" but .. mm:

Kazumi's comments firstly point to the notion that although the views of students in the current ESL context have been accepting of diverse accents in listening assessment to varying degrees, there may be a lower level of acceptance in EFL contexts such as Japan, where exposure to a variety of English accents would be less common than in metropolitan Australia. Her remarks also indicate that test anxiety, triggered by the presence of diverse accented speakers, may be driven by a complex interaction between the very perceptions – increased difficulty, lack of familiarity – that have emerged throughout this study. In this sense, an understanding of test-takers' perceptions is vital in attempting to pre-empt anxiety causing conditions.

6 Summary of emerging issues

The purpose of the focus groups and interviews was not to reach definitive conclusions about the three issues raised in the introduction, but rather to shed light on these through the perceptions of test-takers. We can, therefore, draw from the findings a number of emerging issues related to the research questions. Firstly, "non-standard" accented speech tended to be viewed as a source of additional difficulty for listeners. However, speaker accent was perceived as more salient among higher proficiency test-takers, while lower proficiency learners expressed difficulty with other elements of the task demands. Also, some lower proficiency learners appeared to find the Australian English speaker most difficult to understand, which was at odds with the rest of the participants who tended to single out the Australian English speaker as clear. Secondly, when a shared-L1 accent was noted, it was likely to be

evaluated in a critical light rather than perceived as an advantage in comprehension. The comments surrounding this finding raised a further emerging issue: that test-takers who are distracted by perceived errors in pronunciation and disfluency may lose “belief” in a speaker’s assumed role (e.g., as a lecturer) on a listening test.

Finally, perceptions of the fairness of using diverse accented speakers in listening assessment varied widely across test-takers. Fairness was conceptualised in a number of distinct ways by the test-takers as:

- fitness for purpose: the extent to which the test matched the candidates’ current context
- construct-relevance: the extent to which the comprehension of accented speech was deemed relevant to the test-takers’ communicative needs
- equal access: the possibility for test-takers to gain exposure to, or learn to listen to, an accent
- equal advantage: the potential for an advantage if listener and speaker share an L1
- fair conditions: the creation of reasonable conditions under which to test, specifically avoiding raising test anxiety

It is worth noting that these test-taker-driven concerns match closely the five qualities proposed by Kunnan (2004) in his “Test Fairness framework”: validity, absence of bias, access, administration and social consequences.

The participants in this study who commented on issues of access, equal advantage and fair conditions seemed unified in viewing the introduction of accented speech as potentially violating these aspects of test fairness. However a range of views emerged from the cohort regarding fitness for purpose and construct-relevance which seemed to be based more on participants’ own language learning goals, and their differing perceptions of the sociolinguistic milieu of academic life in Australia.

6.1 Implications

As well as providing some initial insight into test-taker views of accented speech on listening assessment, the findings of this study, and the limitations of the methods employed, hold implications for future empirical research in the area of accent and listening assessment. Firstly, this research shows that learners' stated preference for accents in a classroom setting may not necessarily match their attitudes towards accent in the context of listening assessment. It is therefore potentially problematic to use previous research on learners' accent preferences in an educational environment to predict a likely reaction to the presence of diverse accents in an assessment context.

Secondly, as the findings of the study illustrate, speaker variables apart from accent – such as fluency – varied from speaker-to-speaker, with at least one speaker described as having not practiced enough. It was certainly the case that the oral stimuli used in the pilot study were not of a quality commensurate with a high-stakes standardised assessment: it was obvious that many of the speakers were reading and that sometimes they were encountering, and dealing with, novel lexical items as they read. Thus, while the speakers were representative speakers of particular accents, they were not even in their “performance” on recordings. Future research should aim, if possible, to attempt to select accented speakers of a high level of intelligibility and perceived comprehensibility. This may be accomplished through a more exhaustive rating procedure involving more raters, and a more rigorous methodological approach, possibly based on the methods employed by Munro and Derwing in their studies on accent and intelligibility (see, for example, Munro & Derwing, 1995; Derwing & Munro, 1997). The aim of such a process would be to ensure that the intelligibility and comprehensibility of speakers, regardless of their accent, was shown to approximate that of speakers who may routinely be used for operational EAP listening tests. The more general “believability” of diverse accented

speakers as lecturers, though, would still to a great extent depend on the perceptions of individual listeners during the administration of the test.

Finally, there are grounds for arguing that accent-related difficulty is construct-relevant in an EAP listening test, and this paper has shown that a number of test-takers share this perception. However the crucial issues of fairness related to equal access and equal advantage are valid concerns for test-takers, and clearly there is a need for further empirical investigation to be conducted on whether familiarity with a speaker's accent (or sharing a speaker's accent) may create test bias. Future research should concentrate on these areas as the "weak" links in an argument for the use of diverse accented speakers.

References

Brown, A. & Lumley, T. (1998). Linguistic and cultural norms in language testing: a case study. *Melbourne Papers in Language Testing*, 7, 1: 80-96.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Cambridge ESOL (2008). *IELTS teaching resource*. Retrieved March 1, 2008 from www.cambridgeesol.org/teach/ielts/listening/aboutthepaper/overview.htm

Davies, A., Hamp-Lyons, L. & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, 22, 4: 571-584.

Derwing, T.M. & Munro, M.J. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19: 1-16.

ETS [Educational Testing Service] (2005). *TOEFL iBT Tips: How to prepare for the next generation TOEFL test and communicate with comfort*. Princeton, NJ: ETS.

ETS (2000) *TOEFL® Test Preparation Kit (2nd edition)*. Princeton, NJ: ETS.

Fayer, J.M. & Krasinski, E.K. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37: 313-326.

Fayer, J.M. & Krasinski, E.K. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37: 313-326.

- Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics*, 4, 1: 17-41.
- Hamp-Lyons, L. & Davies, A. (2006). Bias revisited. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 4: 97-108.
- Kachru, B. B. (1986). *The Alchemy of English: the spread, functions and models of non-native Englishes*. Oxford: Pergamon.
- Kunnan, A.J. (2004). Test fairness. In M. Milanovic & C. Weir (eds.) *Studies in Language Testing 18: European Language Testing in a Global Context* (pp. 27-48). Cambridge: Cambridge University Press.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. London; New York: Routledge.
- Lowenberg, P. (1992). Testing English as a world language: Issues in assessing non-native proficiency. In B.B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 108-121). Urbana: University of Illinois Press.
- Lowenberg, P. (1993). Issues of validity in test of English as a world language: whose standards? *World Englishes*, 12, 1: 95-106.
- Lowenberg, P. (2002). Assessing English proficiency in the Expanding Circle. *World Englishes*, 21, 3: 431-435.
- 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.
- Munro, M.J. & Derwing, T.M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 1: 73-97.

Quirk, R. (1990). What is Standard English? In R. Quirk & G. Stein (eds.), *English in Use* (pp.112-115). London: Longman.

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