

## Accents in Listening Assessment

[Start of recorded material]

Host: From the University of Leicester in the United Kingdom, this is Glenn Fulcher with another issue of Language Testing Bytes. In issue 29-2 of Language Testing we publish a paper entitled 'Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective', written by Luke Harding, who is lecturer in Applied Linguistics at the University of Lancaster. This research is concerned with questions surrounding the use of a variety of accents in High Stakes Listening Tests and in this podcast Luke joins us to explore some of the conundrums that face test designers.

Welcome to Language Testing Bytes to talk about your research on the role of accents in listening tests. This is a fascinating field and it's not often that we get papers dealing with the listening construct.

Respondent: Hi Glenn, thanks first of all for inviting me. Yeah, listening does still seem to be an under-researched skill but there have been a number of interesting papers in language testing and in other journals, in the past few years, dealing with listening assessment.

Host: Can we start with a general background question on the topic of accent; you point out in your paper that when language learners go to university they're going to encounter a wide variety of accents but the language tests they take for university entry usually only contain standard native speaker varieties and some people may say that this doesn't reflect the university environment, making the test, in some sense, inauthentic. So why don't the test producers just use a range of different accents on their tests!?

Respondent: Yeah, well, this is the heart of the issue really – why *don't* they? I think there's a few different reasons and they're related to concerns over practicality and validity – so on the practicality side, I think there's a traditional view that L2 speech or second language speech can lead to reduced intelligibility – of course it depends on the speaker and it depends on how proficient they are - but that's one concern I think.

There's also another quite large concern about the acceptability among test takers of using non standard accents or second language accents in listening test materials and there's been a lot of research done on attitudes towards different accents and the research generally suggests that learners tend to prefer native speaker varieties over non native varieties - of course there is research that challenges that as well.

And on the validity side perhaps the biggest concern is the threat of bias, if there is a shared L1 advantage and this becomes an issue of fairness, which is obviously a big concern for test providers. And so I think these

concerns have led to an orthodox approach, where the default position is to use native speaker accents in listening test materials.

Host: You've mentioned the concept of 'bias' here – perhaps we can talk a little bit more about this before we go on – in testing, this has a technical meaning – can you explain that for our listeners?

Respondent: Sure. Well, bias in testing is generally understood as occurring when test takers of an equivalent level of ability but from different groups have a different likelihood of getting an item correct due to some factor which is not relevant to the construct.

Host: Thanks! Now that we've got that straight, can you tell me what your study – which of course you report in your article, in issue 29-2 – was intended to investigate with regard to accent and bias?

Respondent: Right, yeah, well I was interested in following up on this issue of bias and the threat to fairness and specifically I wanted to look at whether listeners who share a speaker's first language would be advantaged when listening to that speaker – so if we take, for example, a list test featuring a speaker with a Japanese accent, would a Japanese listener hearing a Japanese speaker on a listening test, perform better than a listener of the *same* ability who was from a different language background, say a German first language background or a Spanish first language background? – and so I was interested in looking for this shared L1 effect and I used a technique called 'Differential Item Functioning' (or DIF) to look at this in my study.

Host: Differential Item Functioning – now can you be a little bit more specific and tell us what this is and why it's so important?

Respondent: Yeah, DIF is a common way of looking at bias in language testing and in measurement more generally and DIF exists when two groups of test takers who are matched in ability, have a different chance of answering an item correctly. So it's very close to the definition of 'bias' that I gave before but it's not quite the same because a DIF finding may simply mean the two groups differed in a construct relevant way. So it's said that DIF is a necessary condition for bias but it's not a *sufficient* condition.

Host: Okay, I think our listeners will have enough background now to follow your study, so can you briefly tell us what you did and what the main findings were?

Respondent: Yeah. Well I'd used a listening test called the UTESL, which is developed at the Language Testing Research Centre in Melbourne, where I was doing my PhD at the time, and I selected some speakers; three different speakers, to re-record materials for three different versions of the test and I had a speaker with an Australian accent, a speaker with a Japanese accent and a speaker with a Mandarin Chinese accent and they

were selected through a fairly lengthy screening process to ensure that they had equivalent levels of general intelligibility. And I then recruited 212 participants to *listen* to these tests and these participants were from various different language backgrounds, they included 70 listeners from a Mandarin Chinese L1 background and 60 from a Japanese L1 background and all of the listeners took all of the tests and I then used the items in the test scores to look at this shared L1 effect for Chinese L1 listeners and for Japanese L1 listeners.

So, for example, in one of my analyses I matched the ability of Chinese L1 listeners with all the other L1 listeners, using the test which featured the standard Australian English accented speaker and then I performed a DIF analysis on the performance of these two groups, on the test featuring the Chinese accented speaker and I used two different DIF detection methods; the Mantel Haenszel procedure, which is a very common, well known DIF detection procedure and also the standardisation process and I did the same methods in my analysis looking at the Japanese L1 listeners listening to the Japanese speaker as well.

And what I found as a result of these analyses were fairly mixed results actually – the Japanese L12 listeners were advantaged on a small number of items when list to a Japanese accented speaker but this was sort of balanced by DIF against them on other items. And the Chinese listeners showed quite a different pattern, the Chinese L1 listeners were clearly advantaged on at least eight items out of a total thirty-one on the test featuring the Chinese accented speaker.

So there seemed to be a clear effect for the Chinese listeners but at not so clear effect for the Japanese L1 listeners. But this also raised a number of other questions, so obviously one question was ‘why was DIF not observed more strongly with the Japanese listeners when it was with the Chinese listeners?’ but also ‘why wasn’t DIF all pervasive in the Chinese listener analysis? ... why did some items DIF and other items didn't?’.

Host: Okay, well it seems that the results of DIF studies to date, including your own, haven’t provided a clear cut answer to the kinds of issues thrown up by varying accents on listening tests. So perhaps we can finish this podcast by really putting you on the spot; if you were asked to advise and examination board that produces an international English test for university entry on whether they should introduce variety into the accents they use, what would *your* advice be given what we know today?

Respondent: Yeah, well actually this study, as far as I know, was the first application of DIF to this question but other studies have used other methods, usually means comparison procedures, to look at the same issue but the findings have been similarly mixed – so we’ve all ended up in the same place.

But in answer to your question, I think it's a really interesting question actually because it's ultimately a matter of policy ... well, it comes down to a matter of policy at the construct definition stage. And the question is, 'is the ability to deal with L2 varieties (or second language accents) as part of your listening construct?' and if it's not part of the construct then it's clear that the use of L2 accents certainly has the potential to create what would be constructive relevant variants – *but* I think that the arguments for incorporating L2 accents and the ability to deal with L2 accents within the construct really compelling, for reasons of authenticity, modelling the kinds of listening demands that people encounter when dealing with a range of accents.

So the arguments are compelling and I think it's particularly in the case of academic English and to take it out of the construct or to ignore it in the construct definition would mean that you have a sort of impoverished construct I think. So if we take the view that we want it *in* the construct then there needs to be methods for dealing with it and I think that the study that I did, I think one of the implications of it is that it sort of suggests that we can find out *more* about how people deal with different accents and try and understand how we can minimise the impact of DIF and there might be different procedures or different methods that we could use to do this.

So I think, for exam boards or people designing tests, there's a few different ways in which this issue could be dealt with – in the first instance, you could do a needs analysis of different accents in a particular context, so if you were to find out which accents were most common or which accents were necessary to be able to understand in a *particular* situation then any variation in scores would be construct relevant because those accents are part of what is being described in the domain but that's not always possible and certainly very difficult for international tests.

Another option would be to manage the impact of DIF. One of the interesting findings that came out of the study was that the DIF seemed to occur on items where bottom/up processing was tapped and usually items where the correct answer sort of hinged on recognition of a single word and obviously in those cases the pronunciation of that word is all important and if there's any intelligibility problems then that's where the DIF is going to occur.

So one option would be to use item types, where *gist* was being assessed rather than a specifically bottom/up or specific details type questions.

And then another option which has also been suggested by Gary Buck, in his book 'Assessing Listening', would be to use speaker accents which are unfamiliar to everybody, or to use a range of different speakers and balance these over a number of short listening tasks.

So there are different options and there are different methods I think that can be explored to try to deal with this potential for DIF or potential for bias.

Host: Okay. Luke, many thanks for such a candid response and I hope that the representatives of the examination boards are also going to listen to this podcast. Thanks for joining us on Language Testing Bytes and for giving us an insight into this important aspect of listening.

Respondent: Alright, thanks very much Glenn for inviting me.

Host: Thank you for listening to this issue of Language Testing Bytes. Language Testing Bytes is a production of the journal Language Testing, from Sage Publications. You can subscribe to Language Testing Bytes through iTunes or you can download future issues from [ltj.sagepub.com](http://ltj.sagepub.com) or from [languagetesting.info](http://languagetesting.info). So until next time we hope you enjoy the current issue of Language Testing.

[End of recorded material]