



The Electronic Journal for English as a Second Language

Humanising Language Testing

May 2018 – Volume 22, Number 1

M. Obaidul Hamid

The University of Queensland, Australia

<m.hamid@uq.edu.au>

Ngoc T. H. Hoang

The University of Queensland, Australia

<t.hoang6@uq.edu.au>

Abstract

Test-takers' voices in relation to high-stakes language tests have received growing attention in recent years. While the perspectives of this stakeholder group can be utilised to improve test quality, test-taking experience, and test impact, we argue that this goal needs to be achieved by considering a fundamental shift in our conceptualisation of language tests based on test-takers' experiential and perceptual data. This is a shift towards humanising language testing in a globalised world which has witnessed growing expansion, commercialisation, and potential dehumanising of high-stakes tests such as IELTS (International English Language Testing System). Drawing on data from two larger studies, we illustrate the call for humanising the test in seven areas including test purpose, test policy, test-taking experience, test administration, test-taker background, test content and task, and feedback on test performance. Although humanisation has accumulated specific meanings and contentions in the past few decades, test-takers' call for a friendly, responsive, and closer-to-life test and a less stressful test-taking experience may appear more legitimate than controversial in a globalised world.

Introduction

Test-takers of high-stakes English language tests and their attitudes, perceptions and experiences of test-taking and test score use have been of interest in recent years (e.g., Cheng & DeLuca, 2011; O'Sullivan & Green, 2011). This interest underlies the expanding discussion of the role of test-takers in test construction and evaluation, the growing body of empirical work that seeks to integrate test-takers' voices into testing procedures, and the various validation frameworks that have called for incorporating test-taker perspectives (see Hoang, 2018 for details). The recognition of the perspectives of this stakeholder group can be noted in, among various models of test validation, Weir's (2005) socio-cognitive model which considers test-takers as the point of departure, taking into account their physical,

psychological and experiential characteristics. Test-takers can be a valuable source of validity evidence (Bachman & Palmer, 2010; Cheng & DeLuca, 2011), which can potentially benefit the testing product, strengthen decisions based on test-scores and maximise beneficial consequences for all stakeholders. Similarly, Critical Language Testing (CLT), which critiques the power of tests, has argued for introducing people-oriented testing, following humane and democratic principles (Shohamy, 2001). While the perspectives of this stakeholder group can be utilised to improve test quality, test-taking experience, and test impact, we argue that this goal needs to be achieved by considering a fundamental shift in our conceptualisation of language tests based on test-takers' experiential and perceptual data. As indicated by our title, we propose humanising high-stakes language tests such as IELTS (International English Language Testing System) so it aligns more closely with the lives, backgrounds, expectations, and experiences of test-takers. We are aware that the trade-off between ensuring fairness in the application of testing standards (*equality*) and promoting *equity* between test-takers (Lam, 1995) cannot be ignored. Nevertheless, we would assert that in the apparently divergent and heterogeneous voices of test-takers, it is possible to extract a coherent narrative that gives greater weight to the first option from the "either/or" proposition mentioned by Brown (2004):

What should drive test design? Should it be characteristics of the people taking the test, or should it be the purpose of the test and the decisions being made with it? (p. 319)

If the growing interest in test-takers and their perspectives is motivated by a desire to respect and act on their voices, we may need to consider them not just as test-taking subjects but as human beings taking language tests in the pursuit of human desires in a globalised world. We seek to corroborate our argument by drawing on two studies that focused on voices of test-takers of the IELTS test. In what follows we first present our conceptualisation of humanising, drawing on the language teaching and testing literature. We then introduce the IELTS test as a gatekeeper in a globalised world. This is followed by our presentation of the two studies together with data analysis. A brief discussion of the findings is followed by our conclusions at the end.

Humanising in Language Teaching and Testing

Our understanding of humanising can be traced to several movements in language teaching and testing in the post-1970s era. In terms of the former, humanising has had somewhat of a controversial history which can be understood from a number of articles published in *ELT Journal* (e.g., Arnold, 1998; Atkinson, 1989; Gadd, 1998). Humanistic language teaching, which was debated in these writings, focused on humanistic approaches to second language (L2) pedagogy associated with the humanistic psychology of Carl Rogers and Abraham Maslow underpinning the Silent Way, Suggestopedia and Community Language Learning (see Richards & Rodgers, 2001). Key characteristics of these approaches include: a focus on the whole person, nurturing of the self and motivation for self-realisation, affective well-being, and emphasis on subjective experience. This version of the humanistic project was critiqued for its: a) romantic rather than pragmatic humanism; b) introspection at the expense of knowing the world; and c) affective development at the expense of cognitive development

(Gadd, 1998). Although misleading simplification of humanistic language teaching underlies many of the critiques (see Arnold, 1998), humanistic approaches “accumulated a certain amount of problematic baggage” by the late 1990s (Kerr, 2007, p. 5).

A more recent reference to humanistic language education is the journal *Humanising Language Teaching* published by Pilgrims, which has deliberately distanced itself from the above-mentioned humanistic psychology to signify more general meanings (Kerr, 2007). These include humanising in the sense of positive and desirable as opposed to dehumanising, which is negative and undesirable. As such, humanising is comparable to typical dictionary definitions of the term such as “rendering human or humane”, “imparting human qualities”, and “making a process or system less severe or easier for people to understand”.¹ Thus, “humanistic” appears to be a synonym for “good”; “humanising” meaning improving or making things better: “how we have made or could make things better in our classrooms” (Kerr, 2007, p. 6).

The other provenance of humanising can be located in the communicative language testing of the 1970s and early 1980s. This emerged as a reaction against multiple choice tests (see Hoffmann, 1962) which had an overemphasis on reliability (Fulcher, 2000; Morrow, 1979). The communicative movement emphasised humane qualities, as it was seen as a

rebellion against a perceived lack of concern for the individual, the human, the “subject” taking the test; it wished to remove the use of so-called “arcane pseudo-scientific jargon”, and rejoice in the common sense of the classroom teacher who can tell whether a test is good or not by looking at it; it wished to “re-humanise” assessment. (Fulcher, 2000, p. 484)

The “re-humanising” movement highlighted: 1) real life tasks; 2) face validity; 3) authenticity; and 4) performance (Fulcher, 2000, p. 484). While these characteristics resonate with our conceptualisation of humanising, our data also allow us to go beyond the parameters of communicative testing. This may be because while communicative testing originated in the thinking of language educators, our conceptualisation is based on test-takers’ perspectives. In our grounded view, we can see traces of humanistic approaches, communicative and critical language testing as well as humanistic education in general, as championed by Paolo Freire and applied in language teaching by scholars including Akbari (2008) and Canagarajah (1999) as well as more general meanings of the term found in dictionaries, as previously noted. We can highlight the following features of humanising in our conceptualisation in language testing:

- making tests meaningful to test-takers,
- making tests life-like,
- drawing on test-taker backgrounds and their needs and concerns,
- imparting human qualities to the system of testing,
- making test-taking experience pleasant and stress-free, and
- improving test qualities.

Our conceptualisation strongly aligns with human and social conditions in educational assessment which have underscored the human experience in assessment (Harris & Brown,

2016). As these authors elaborate on in their introduction to a recent handbook by the same title:

Hence, when we speak of human conditions, we are talking specifically about emotions, experiences, and beliefs which occur WITHIN an individual and influence how that person understands, engages with, and interprets assessment experiences and results. (Harris & Brown, 2016, p. 2, highlight in the original)

Why humanise high-stakes language tests?

Our call for humanising IELTS along the lines of the above-mentioned conceptualisation should be considered against the global expansion and increasing commercialisation (and potential dehumanising) of English language tests as massive global systems (Hamid, 2016). The dehumanising potential of language or education tests cannot be denied. Tests require test-takers to submit to their “invisible power” physically, emotionally, and intellectually. This was clearly illustrated by Peirce and Stein (1995) in a reading test validation study that involved African students in a school setting in South Africa. The authors reported a “carnavalesque” moment once the dehumanising spell of testing was over and the test-takers returned to their “normal” life. Although probably a coincidence, the test-takers also thought that the monkeys that they had read about in the reading test stood for themselves—the identification evoking the whole socio-political history of white rule in black Africa.

The potential of dehumanising is higher for high-stakes tests, which treat people as test-taking subjects often with little consideration of their human needs, backgrounds and experiences, although, arguably, in the interests of fairness and equality. Not being able to accommodate test-takers’ needs, backgrounds and experiences may be a systemic issue, which itself may justify injecting human qualities into IELTS which has emerged as a globally operative “mega system” of testing (see the next section).

Humanising language tests or taking human factors into account is also a requirement for test validity. Referring to Messick’s (1984) emphasis on the personal and environmental factors, Harris and Brown (2016) argue that understanding human and social aspects of assessment is “important for the design of assessments, the preparation and development of teachers, the quality of scoring and marking systems, the creation of appropriate policies, and the design of statistical models used to generate scores” (p. 3). Nevertheless, as these authors argue, human and social factors have received minimal attention in research.

Finally, the use of language test-scores for life changing decisions has serious consequences for some test-takers. These consequences may call for understanding, from a human perspective, how test results affect people’s lives. Although we know little about how global English tests affect the lives and dreams of test-takers, the few cases reported (e.g., Ahearn, 2009; Hoang & Hamid, 2016) would justify our call for humanising. For instance, the following excerpt is taken from the Radio National website in Australia, which hosted popular comments on the use of compulsory English test for immigration.

Really you can't sleep... When you're sleeping you are dreaming about IELTS and when you eat, you walk, whatever you do, you are just stressed about everything around your world. And I would say, like, you know, it's really a do or die situation.²

This experiential account indicates the extent to which the stakes involved in tests affect the lives of some people taking high-stakes tests.

IELTS as a Gatekeeper in a Globalising World

IELTS is jointly owned by British Council, IDP: IELTS Australia³ and Cambridge English Language Assessment. It tests the English proficiency of L2 speakers in listening, reading, writing and speaking. Test performance is measured on a 9-point scale in which band 1.0 denotes "non user" and 9.0 stands for "expert user". Test-takers' raw scores in the four components are calculated into an aggregated band score on the 9-point scale for each test-taker. The test was introduced in 1989 and since then it has undergone several revisions based on in-house research on various aspects of the test, test administration, and impact (see Chalhoub-Deville & Turner, 2000; Stoyhoff, 2009).

The IELTS website⁴ claims that it "is the world's most popular English language test" for study, work and migration. In 2016, over 2.7 million tests were taken in 1,100 locations in over 140 countries. The test's acceptability has been extended globally with the number of institutional test-users exceeding 9,000, including schools, universities, employers, immigration authorities and professional bodies in both traditional and emerging English-speaking polities. Educational institutions and immigration departments in these countries have set up different IELTS score requirements for prospective students and/or visa applicants.

The global expansion of IELTS may help enhance its technical quality and popularity by utilising its accumulated resources and expertise. However, continued expansion may also affect its educational identity, reinforcing its business identity (Davidson, 1993; Templer, 2004). While educational commodification is a global concern (Luke, 2004), some questions for IELTS and similar tests concern how their profit-making agenda aligns with the stated goals of measuring language proficiency and whether tests driven by profit-making goals pose social, educational and ethical issues and challenges (see Davies, 1997; Sarich, 2012). It is against this growing use of IELTS as a gatekeeper of global flows of people on the one hand, and its growing commercialisation and profit-maximisation on the other that the present article examines voices of test-takers drawing on two larger studies.

The Two Studies

As previously noted, the data for the present article were drawn from two studies which dealt with voices of test-takers of high-stakes language tests. Ethical clearance was obtained for both studies from two Australian universities.

The first study focused on test-taker perspectives on IELTS in a globalised world (see Hamid, 2014 for details). The main instrument for data collection was a survey which was completed

by 430 test-takers (female=45%, male=55%) from 52 countries. This was followed by interviews and a focus group discussion with test-takers recruited from the survey sample, taking into consideration the diversity of the sample and the general patterns in survey responses. The questionnaire was divided into three sections. The first section asked for personal information from test-takers in relation to taking the test (i.e., for what purpose, how many times and what scores). The second section contained 40 items arranged on a 5-point Likert scale which focused on aspects of the test, test-taking experience, and test use. The third section included an open-ended question that sought their feedback on the test and suggestions for further improvement of the test. This non-obligatory open-ended question was completed by 343 respondents (80%). These responses together with the interview and focus group data were drawn on in the present article.

Although there were four items related to test-takers' experiences (i.e., test-taking as a pleasant experience; that it caused stress and anxiety; that it encouraged improving English; the feeling of helpless) of taking the test, the questionnaire had no explicit reference to humanising. The themes that emerged from the open-ended comments and were articulated by the test-takers in many ways in relation to the following main aspects of the test, test administration and reporting were:

- Test purpose
- Test policy
- Test-taker background
- Test content and task
- Test-taking experience
- Test administration
- Feedback on test performance

The second study explored test-takers' perceptions and experiences of IELTS and TOEFL (Test of English as a Foreign Language) with the aim to theorise their contribution to the validation of high-stakes testing. A test-taker validation model featuring four steps—domain definition, evaluation, extrapolation, and utilisation—adapted from Chapelle et al. (2008), informed data collection and analysis. In the first phase, 518 participants (377 IELTS and 141 TOEFL test-takers) completed an online survey and 260 of them also provided open comments. The survey focused on 1) test-takers' demographic details; 2) their experiences with the tests; and 3) their perceptions of the tests' reliability and use of test scores. In the second phase, 28 survey respondents participated in semi-structured interviews while another 10 test-takers participated in three focus groups.

The qualitative data drawn from open responses and individual and focus group interviews were examined for instances of humanising in the seven areas identified from the first study. In the next section we illustrate humanising by referring to data from both studies. Participant codes that start with R refer to the first study while those starting with S refer to the second study.

Findings

Test Purpose

When a language test score is required as a proof of *basic* language proficiency, test-takers who considered themselves proficient users of English and/or had other evidence of language proficiency (e.g., a score in equivalent tests, or work experience) tended to question the rationality of this requirement. They had the perception that the requirement was a bureaucratic imposition, which was meant to enable score-receiving institutions' quick, easy, and cost-effective processing of applications by leaving the hassle and associated costs to test-takers. Therefore, even when test-taking experience is represented as positive, these test-takers perceived the requirement as unnecessary and unjustified. As RJ (interview participant) pointed out: "yes, it [taking the test] was good experience. Personally I like it. But what makes it not interesting is that it is a condition to fulfil." To him, it was just a bureaucratic necessity; he could not make sense of the requirement in any other way. R144 (survey participant) did not like the test, and did not see any point in taking it:

I really don't like the test, but I have to take it because of work and immigration. I studied and worked in AUS [Australia] for 8 years, there is no problem for me to live and work here...

S209, an interview participant from India, observed: "I know that this is going to be a waste of my time and money as well, because i know how well i can speak English and write in English." He explained this perception by referring to his language background on the one hand and the cost involved in preparing for the test on the other:

English is my first language, i had been learning in English since i started school and everything [...] all the tests and exams that i ever had in my life were in English. In India, almost all the most competitive exams, 98% of them, are held in English. You know, i had to prepare not only English but also the way the test assesses it, i knew that it was not just only speaking, writing, listening, reading, but i had to prepare properly the skills that were required to do the test. 9,000 rupees is not a small amount, it has to be considered a big amount. You know, it's just an exam but it costs 9,000, it's very expensive.

Similarly, S81, whose score of 8.5 out of 9.0 in IELTS expired three weeks before applying for an MBA course at an Australian institution, found it inappropriate for the institution to require her to provide a valid score of at least 6.5. Like S209, she saw this as a waste of money and time for her. As she argued:

[I was] a little bit annoyed, because you can't go from 8.5 to 6.5 in three weeks' time, right? I mean, from 6.5 to 6.0 is possible, that's definitely a dangerous result. But when you've got beyond 7, it's hard to go back to a really low score because you've already got the foundation, skills, and vocabulary, etc. Apart from that, I was an English teacher for seven years, and I was a lecturer in an international leading university... well, but this is the policy. That's also a reason why I think you should always look at the *person*, not the test score. (emphasis added)

There is a clear call for person-focus (read humanising) in the voice of this test-taker. The test-takers would have preferred a case-by-case consideration of individuals based not solely on the test score but also other legitimate evidence of proficiency. As S81 maintained:

I believe that test result is just a number and even if this number makes an indication of that person's specific knowledge of English at a certain time, overall I still think that the whole person, the whole process is a more important indicator than a test paper or a number.

There is also an indication that if people studied and worked in an English-speaking environment for a substantial period of time, they should be exempted from taking the test. Among those who most strongly supported this more flexible policy is a Peruvian test-taker (S185) who had lived in Australia for more than eight years, had Australian qualifications, and got a decent job as an accountant. However, she had to leave Australia because she could not satisfy the IELTS score requirements for permanent residency visa application. As she explained:

For this IELTS, I have to go back home. And at home, I will have to start again, but it's different because *it has changed your life*. When I came here I was eight years younger but when I go home, I am eight years older. People always say that money is what you can make, but the time once gone it never comes back to you. So I think that *IELTS took my life!* [...] When I first came to Australia, I was full of hope, full of energy and youth, without grey hairs. All my energy has been squeezed and IELTS has taken my money and my time. (emphases added)

It is relevant to note here that a Dutch-speaking academic at Australian National University felt insulted when he was asked to take the IELTS test by the immigration department for Australia (van der Heijden, 2013).

Test Policy

Test-takers were critical of certain policy aspects of IELTS which, in their view, were neither justified nor congruent with their interests. Specifically, they referred to two policies: the test retake policy and the score use policy. In terms of the former, they suggested that instead of asking test-takers to repeat all four components of the test, they should be allowed to repeat only those components in which they had not obtained their required scores. As R362 pointed out: "Whichever area the test taker had a lower band score should only be the area to re-take". Similarly, R67 noted:

If someone fails to achieve a good score in one section of IELTS test, he or she should give [sic] an opportunity to re-appear on that section only.

This view resonates strongly with a number of other test-takers who had to take the whole test multiple times to satisfy specific sub-score requirements. In cases where test-takers repeatedly missed the target score by a narrow margin of 0.5, the question of the limit of the language test was frequently raised. Having taken the IELTS test four times, S370 (interview participant) posited:

My greatest impression about the test is that even if your ability equals to 7, you don't necessarily always get 7 on the test. What happens is that you actually fluctuate around it several times before you can get the score that reflects your real ability. This is because you

are very likely to make mistakes during the test and any minor mistake can drag your score down. For example, in your listening test, a single inattentive moment can cost you several answers, which will be converted to a band score off your real mark. From my experience, a person of 7.0 level generally needs to take the test at least 3 times to actually get a 7.

S375, an interview participant, failed to fulfil the requirement of 7 after 14 attempts, although he was able to obtain the required score for each of the test components across the test sittings. As he pointed out:

I think IELTS and TOEFL are among the most reliable language tests being used in the present time. But no test can be perfect. They're essentially assessing only a sample of your whole range of language use. So the rules of probability apply.

Alternatively, the test-takers suggested that multiple scores on the IELTS test taken across time to satisfy score requirements for individual components should be accepted by institutions and authorities using the test. As R281 pointed out:

If a candidate attains several tests then it should consider highest score in each band [band] and overall band should be considered on the bases of those highest scores.

Test-takers felt that requiring them to satisfy overall and skill-specific score requirements in a single sitting of the test was too restrictive given that there were significant variations in scores across time (see Hamid, 2016). They observed that their suggestions were reasonable given that scores remained valid for two years. As S376 (interview participant) pointed out:

IELTS itself says that the test score is valid for two years. But in fact, the way the score is required means that the score is invalid the moment you get a new score. For example, if you got 8.0 a month ago and now you take it again and get 6.5, they will only take your score as 8.0. It contradicts IELTS's own statement. If not, why don't they accept my combined results?

Test-Taker Background

Test-takers called for IELTS to be made "more test-taker friendly" (R171) which can be done, it is suggested, by taking into account their backgrounds. They felt that currently the test did not fully reflect their interests or situations. For example, R24 explained:

...I feel that the topics, themes and situations chosen for different sections of IELTS are unnecessarily, yet strictly 'Western' which often challenge test-takers' cultural awareness rather than language skills!

In particular, S296, a Muslim test-taker from Saudi Arabia, indicated a cultural barrier to his test preparation and performance:

[...] you [will] study only in your own field, using English only in your subject, then why do you have to learn different cultures, different lives, different artists for example [...]

example, in my culture it's not appropriate to listen to Rap for example, or music. But you have to [...] Sometimes they ask you very silly questions that you won't even think about. [They] push me into pressure to learn about house warming for example, or about different foreign [rituals] that are really strange.

Similarly, S81, a Romanian test-taker of both IELTS and TOEFL commented:

I was born and grew up in Europe and a lot of the topics that I saw in the [IELTS] writing task 2 sounded like science fictions to me ... like "Who cares about this??? Why do I have to write about it?" ... I don't know, but I can't find anything relevant, I can't find something I can relate to.

Despite issues related to the cultural orientation of the test, the high stakes associated with the test result left test-takers with no alternative.

As S296 stated, "they put a lot of pressure, they apply and impose these things on us. We have to do it otherwise we'll be kicked out of the country."

Test-takers expected the test to be more relevant to their socio-cultural and language backgrounds. For example, R117 noted: "I do believe that readings and writings should be based on the test-takers' purpose and background". Similarly, R34 suggested devising such procedures which would ensure ease and pleasure in taking the test for test-takers who are non-native speakers of English:

For IELTS, such a test procedure and test language should be chosen where the non native speakers can participate with ease and pleasure. It will be familiar to them both linguistically and culturally.

Test-Taking Experience

The two studies documented candidates' mixed experiences of taking IELTS, while there was a call for making test-taking a more pleasant experience. As R12 wrote: "To me, honestly, I do not like IELTS, because it costs me a lot and gives me a large scale of stress and pressure". RC (interview participant) conveyed a note of pessimism by saying that there was "nothing happy, nothing good, everything is bad, a very bad experience". R99 reported that she "fainted" during the test, but she "was not given extra minutes, not even 15 minutes". While the underlying reason for not giving extra time can be appreciated by referring to the standardisation of procedures to ensure that everyone takes the test under comparable circumstances, it may be necessary to think about individual needs. Test-takers suggested that the test should be friendlier which can reduce stress and anxiety. As R140 pointed out:

The test designers should think of how to design a more friendly test that helps to reduce stress and anxiety for test-takers.

Test-takers' experiences of taking the test were linked to the possible consequences of failing to achieve the desired score, specifically by referring to the opportunity and financial costs involved. For example, S13 explained:

So generally talking about my IELTS experience, i was very nervous because of the huge investment [...] in comparison with your salary and because of the high stakes it has, when you apply for the scholarship and have to prove that you are able to get the score.

RN, S107, S221 and S376 also attributed their unpleasant test experience to the high fee of test registration, which, according to S107, was “intolerably exorbitant”. S221 suggested that the test provider (IELTS) should find out ways to reduce the test registration fee to make the test more affordable and accessible to a larger number of students which can also contribute to a more pleasant test-taking experience.

Test Administration

While adequate attention is given to the administration of IELTS from the point of view of test security, reliability and fairness, aspects of the “bureaucratic process” (RJ) appear not to take the basic needs of test-takers into account. R353 observed that IELTS “test takers should be given a break after the listening and reading test”. The break would provide an opportunity to test-takers to go to the toilet, as pointed out by RN, R218, R353 and R386. As RN stated: “However, they even have to refrain from going to the toilet otherwise they will lose precious time for the subsequent tests”. S185 considered that not being able to go to the toilet was harmful for test-takers’ physical health:

I never drink water during the four hours of the test. I also have to stop drinking water an hour before that. If you think about this, it is also against your HEALTH! Your body needs water to be healthy, and you need to go to the toilet, but no, you can’t drink it during the test.

More critically, it was pointed out that IELTS “invigilators were behaving roughly with the students” (R55), who “treat test takers like potential cheaters [sic] with no morals” (R364).

A perception held by a large number of test-takers was that the administrative procedures were aimed at identifying frauds but in doing so, they disregarded the self-esteem of “honest” test-takers. S185 described the procedures in the following way:

It’s a very strict process, maybe because they think that you’re gonna cheat, even if you need to go to the toilet, there will be people to follow you. You have to leave everything outside, you have to leave your mobile phone, they also check your pen and eraser. All in all, it really feels like you’re going for a torture!

Additional anxiety and a lack of respect were experienced by S296 who commented:

The test pressure is there because they follow the same procedure, the same way of dealing with you, you know, finger printing, checking your passport, nodding, no smiling, saying nothing, they just deal with you as guilty or a criminal. Yes, very much like that. Reminds me of the police. You can’t imagine that.

Concerned about the accommodation of disabled test-takers, S221 reported what he considered rigid and unreasonable at an IELTS test centre in Pakistan. According to him,

while his requests to have a special device for test-takers with hearing impairment and preferred speaking test time were approved by one test centre, these requests were not given timely attention in another centre.

The arrangement of the speaking section of the IELTS test on a subsequent day rather than on the same day with the other three sections of the test were also not favoured by the test-takers due to the tiring wait, prolonged test anxiety or being suspended “in-the-test kind of feeling” (S80) and extra accommodation costs for test-takers who were not within daily travel distance to the test centre (S221).

Although test-takers were generally aware of the rationales for the foregoing procedures, they expressed their wish for the procedures to be made more test-taker-friendly. High-stakes tests are characterised by test anxiety and these procedures, they noted, should not create excessive anxiety that may interfere with test performance.

Content and Task

In addition to aspects of test-administration, test-takers called for changes in some aspects of the test content together with procedures for task performance. An overriding concern was the use of test prompts and tasks that did not reflect the lives and experiences of test-takers. Space does not permit providing detailed examination of these internal aspects of the test, but we cite a few examples from listening and speaking components of IELTS.

Many test-takers mentioned the True/False/Not Given question type in IELTS, which they thought were “very tricky” and more like a guessing game than a test of reading comprehension (S40, S80, S248 and S291).

In terms of listening, RJ pointed out how IELTS listening did not reflect listening in real life. As he explained: “When you speak to somebody, if you can’t hear well, you’ll say ‘Pardon, I can’t hear you’, then the person will repeat. But the test’s only once.” While those involved in testing and tests may understand that this compromise is a necessity in a test environment, test-takers may not have the assessment literacy (Taylor, 2013) to appreciate this principled compromise. However, instead of acknowledging the limits of testing and testability, the IELTS authorities seem to emphasise the life-likeness of IELTS in representing the test on their website. This may encourage test-takers to look for evidence from the test that problematises claims of the test’s connection to real-world language use.

Speaking, in their view, appears to be unrelated to real-life speaking. As RD pointed out: “But as far as I can see up to now, it’s only one-way. It’s not really a conversation. It’s totally different from what you’re doing in the daily life.” Apart from the nature of the speaking tasks, the overall gesture and behaviour of the speaking examiner may also contribute to making the speaking test artificial. As R262 noted: “The speaking test is not very interactive. The interviewers appeared to be robotic”. The test-takers’ views provide an approximate characterisation of the role of the interviewer who simply asks questions, rendering the linguistic exchange a kind of discourse that cannot be found in real life. Although some of the issues raised by the test-takers may challenge test designers to define the boundary of test

situation and language use in real life, some of their suggestions are practical which can make the test more life-like and test-taker friendly. For example, as R171 suggested:

Also, before the recording begins, in the speaking test, a warm-up discussion session would be helpful where the interviewer would ideally tell the candidate not to take the test as an interview but rather as a friendly discussion and they can also ask question. Too often candidates tend to get too tensed and their interpersonal and pragmatic skills remain underused.

It appears that the issues raised by test-takers have the potential to affect test-takers' performance which may point to validity issues. As R63 observed:

The organiser should try to make the interviewing process as friendly as possible as sometimes, the reason people don't do well in interview is because they are nervous.

Test-takers suggested that the speaking "should be a conversation not a[n] exam" (R262); it should be "friendly" (R63, R171), "casual" (R76) and "conducted in a relaxed manner" (R376). These requests can be related to the inherently social character of language performance in everyday life and in the testing context (McNamara, 2001). If test-takers perceive the contribution of the speaking examiner as less facilitating, then there might be questions of validity.

A more critical issue in terms of speaking appears to be the choice of topics for the speaking tasks. While it is generally understood that topics and tasks should reflect the target language use domain, the choice of topics also raises the question of whether language ability needs to be seen as being embedded in the topic or whether this ability can be de-coupled from the latter. Test-takers pointed out that they often encountered topics which were unfamiliar to them which affected their performance. As R133 explained: "we can't talk much about the topics because we don't know much about them, not because we can't speak English well." R231 and R232 provided examples of topics such as "dance" and "music", about which not many people in their country (Saudi Arabia), they asserted, would know. They could not speak much about these topics to demonstrate their speaking ability because, as R332 noted, "it is very hard to speak about subject you just don't care". RS provided a different interpretation of speaking topics which are unrelated to test-takers' lives. As he noted:

To me the speaking test is encouraging people to lie, to laugh or to do non-sense in order to impress the examiners so that they can say that this person who's taking the test is able to do so much work, explain a lot of difficult things. In some sense, the question in the speaking test is sometimes really silly so you don't have any options rather than doing some make-up thing, which is pretty much lying.

One may not agree with this view, but there may be some logic: If a test-taker is asked to talk about their favourite something, even if she/he does not have one, s/he has no choice but to make up something which, to this particular test-taker, appeared as good as lying. A number of other participants reported having "lied" in the speaking test.

S376 related the test topic to issues concerning the consistency of test difficulty:

If test-takers are native speakers, they may find IELTS topics as of similar difficulty levels. But the majority of IELTS test-takers are non-native English speakers, and to them, definitely some topics are more difficult than others. This is why I think taking the test is a matter of luck. If you're lucky, you can get a topic that you're familiar with or are interested in and get a good mark. If you're unlucky, you get something that you have no idea about, you get stuck and can't show your English knowledge and skills.

Research into the extent to which test content and task allow test-takers to demonstrate their language proficiency deserves attention, especially because most of the test-takers believe that topical knowledge was an important factor in test performance (see Elder, Iwashita & McNamara, 2002).

Rater Feedback

The last example of humanising IELTS is related to the absence of feedback on test performance. Given that IELTS is taken by a large portion of test-repeaters (Hamid, 2016), these test-takers believed that they would benefit from feedback on their performance. However, IELTS does not provide detailed feedback.⁵ Therefore, test-takers pointed out this absence. As R180 said:

The test should be transparent with feedbacks to improve ones English [sic]. Since it is a paid test, the payer should have right to know his mistakes before taking another test.

Similarly, R384 observed that "Feedback should be given so that test takers can understand from their mistakes". R180 hoped that IELTS will have feedback system in the future which will "further help test-takers to improve". Importantly, feedback is not available even from the result review process for which test-takers have to pay extra fees. As R387 explained:

Moreover, the appeal of IELTS results was very expensive and unfair because it does not let you know exactly where did you make the mistake. Someone will just inform you that a review of your test has been done and there will be no change to the IELTS band score since the results given was reasonably fair and correct.

The non-provision of feedback is probably motivated by the necessity of maintaining security and confidentiality of the test. It might also be consistent with the role of the test as a proficiency test rather than an achievement test. However, test-takers tended to perceive this absence of test feedback and the enquiry-on-result process negatively. S147 questioned the transparency of the test while S376 speculated:

I think that they intentionally make requesting a re-marking of test paper difficult so they to discourage people from doing so to make test administration easy and cheaper [...] You have to think hard because the fee for that is also very high and it takes very long to process the request, two to three months. And it's risky because you may lose time and opportunity. You

will also lose money if the score isn't changed. And the wait for the result is terribly stressful. So it seems to encourage people to sit the test again because the test is now quite frequent.

On the other hand, S185, S352 and S376 believed that this policy was driven by financial motives: "They want us to sit the test again and again and again so they can earn more money" (S352). While test security is important for all stakeholders, feedback is probably a legitimate demand, at least this is how test-takers perceived it.

Discussion

In this article we have discussed test-takers' perspectives on the IELTS test focusing on seven areas including test purpose, test policy, test-taking experience, test administration, test-taker background, test content and task, and feedback on test performance. These are overlapping and interdependent categories which can be divided into two groups: categories related to the test content itself, and those to the conditions under which the test is given. We have presented them separately for convenience of analysis and reporting. Test-takers' views of these areas suggest that they are calling for making the test more responsive to their realities, needs and concerns. They are urging authorities to bring the test closer to their lives— hence humanising the test. Based on test-takers' voices represented in the paper, we can respond to Brown's question cited earlier by pointing out that test design should preferably be informed by test-taker characteristics. Thus, we would argue that if test-takers are given an opportunity to make their voices heard, which has happened increasingly in recent years with the increase in test-taker research, they would probably urge testing authorities to humanise the test.

How important are the issues raised by the test-takers and how appropriate is the call for humanising IELTS? Answers to these questions can be provided from different perspectives. In terms of the aims and principles of language testing, it appears that some of the issues may have affected test performance. If tests and testing conditions are expected to create an optimal environment in which test-takers can demonstrate their best language ability, it can be pointed out that there may have been some compromise in this regard. To the extent that this has happened, there can be questions of validity of IELTS—that is, the test not capturing test-takers' ability in its entirety. Listening to the test-takers is also important for consequential validity—that is, how high-stakes tests affect test-takers' lives and life chances, as we have previously argued.

Assuming testing authorities are willing to give reception to test-takers' voices, is it possible to take all their suggestions on board? Presumably not. This is because some of the issues test-takers raised cannot be addressed within the limits of language testing (consider the case of not having the opportunity to ask for repetition in the listening test, as pointed out by a test-taker). However, a humanist approach may be useful here which may work in favour of testing agencies in pre-empting some potential criticisms. In emphasising the technical excellence of IELTS together with its reliability, authorities may acknowledge the limits of testability. For example, an acknowledgement that language or education tests given under test conditions may not fully capture the complexity of language use, or that there is always a gap between scores obtained under test conditions and language use in real life, should help authorities to win the trust and confidence of test-takers.

Moreover, we acknowledge that there were contradictions in test-takers' views—while many test-takers talked about making the test more closely aligned to test-takers' lives and experiences taking an equity perspective, some also supported what is called procedural fairness, being informed by fairness and equality perspective (see Hamid, 2014). Although we haven't reported their views of equality (discussed in detail in Hamid, 2014), we need to underscore that test-takers' views are much more complex, underpinned by both issues of equality and equity. At the same time, we also acknowledge that in the data there were voices, which constitute a minority, that do not consider it important to focus on test-taker backgrounds. As R310 observed:

The process of testing in it's [sic] essence is meant to be judgemental, not politically correct and non native-friendly. It is to be English-friendly, it tests English language proficiency.

Nevertheless, the majority supported the view of making the test more humane, that is, incorporating human considerations into the test design, test content and test administration and improving the test so that test-takers have more positive experiences and they can demonstrate their best performance. It appears that bypassing the “either/or” dichotomy of the equality-equity debate it is possible to implement some of the feedback provided by test-takers. For instance, even without changing the format of the IELTS speaking test, examiners may be able to show more warmth and welcoming gestures and establish good rapport with the interviewees. It may also be possible to make the test more interactive so the examiner appears more human and less robotic. Similarly, invigilators can demonstrate professionalism and treat test-takers with respect and care. Test centres need to welcome test-takers with special needs and accommodate those needs. A short break of 10-15 minutes after listening and reading may not seriously affect the test administration schedule but it would be highly appreciated by test-takers. Finally, giving feedback on their performance on writing and speaking appears to be a legitimate demand and testing authorities may explore the ways and means of implementing this demand.

Conclusion

As a global leader of English language testing, IELTS is one of the most widely used L2 tests of current times. The test has marked a radical departure from the “pseudo-scientific jargon” of mechanised tests previously mentioned. Although IELTS is represented as close to real-life language and language use in terms of the target domains of language use (see the IELTS website), it is probably not close enough to the lives, situations, and expectations of the diverse group of test-takers taking the test. With the growing recognition of test-takers' voices, it may be possible to listen to the test-takers behind the language data and humanise the test. Given the growing expansion and commercialisation of IELTS, it is probably imperative for test authorities to imbue the test with human qualities, at least to the extent this is feasible within the limits of testing and testability.

The theorising of test-takers' perspectives proffered in this article has emerged from test-taker data. However, we understand that the data may have limitations in that these were contributed largely by test-takers who may have had less pleasant experience with the test. Further research involving larger samples of test-takers with both positive and negative test-

taking experiences will be useful to verify the call for humanising. This research may also draw on test-takers of other global English tests.

Endnotes

1. See, for example, the Oxford and Merriam-Webster dictionaries. [[back](#)]
2. Please refer to <http://www.abc.net.au/radionational/programs/backgroundbriefing/language-barriers/2948964> [[back](#)]
3. International Development Program of Australian Universities and Colleges Ltd (IDP) is owned by 38 Australian universities and the job website SEEK. It provides international student placement services. See www.australia.idp.com [[back](#)]
4. www.ielts.org [[back](#)]
5. IELTS has recently introduced a new service which allows test-takers to receive expert advice on IELTS speaking, reading and writing skills. For instance, prospective test-takers can perform a sample writing task and receive feedback by paying a fee of \$50. Please see <https://ielts.com.au/study-for-ielts/exam-help/>
While some test-takers may benefit from this service, this is not what the participants in our studies were referring to. [[back](#)]

About the Authors

M. Obaidul Hamid is Senior Lecturer in TESOL Education at the University of Queensland, Australia. Previously he worked at the University of Dhaka, Bangladesh. His research focuses on the policy and practice of TESOL education in developing societies. He is the co-editor of *Language planning for medium of instruction in Asia* (Routledge, 2014). He is on the editorial boards of *Current Issues in Language Planning*, *English Teaching Practice & Critique*, and *Journal of Asia TEFL*.

Ngoc T. H. Hoang has just completed her PhD thesis at the University of Queensland, where she is currently involved in teaching and research activities. She has a background in language education and assessment. Her research interests include the social aspect of language testing and assessment, test-takers' voices, validity and validation.

References

- Ahearn, S. (2009). "Like cars of breakfast cereal": IELTS and the trade in education and immigration. *TESOL in Context*, 19(1), 39-51.
- Akbari, R. (2008). Transforming lives: Introducing critical pedagogy into ELT classrooms. *ELT Journal*, 62(3), 276-283.
- Arnold, J. (1998). Towards more humanistic English teaching. *ELT Journal*, 52(3), 235-242.
- Atkinson, D. (1989). 'Humanistic' approaches in the adult classroom: An affective reaction. *ELT Journal*, 43(4), 268-273.

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: developing language assessments and justifying their use in the real world*. Oxford; New York: Oxford University Press.
- Brown, J. D. (2004). What do we mean by bias, Englishes, Englishes in testing, and English language proficiency? *World Englishes*, 23(2), 317-319.
- Canagarajah, A. S. (1999). *Resisting linguistic imperialism in English teaching*. Oxford, UK: Oxford University Press.
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28, 523-539.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.
- Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, 16(2), 104-122.
- Davidson, F. (1993). Testing across cultures: Summary and comments. *World Englishes*, 12(1), 113-125.
- Davies, A. (1997). Introduction: The limits of ethics in language testing. *Language Testing*, 14(3), 235-241.
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing*, 19(4), 347-368.
- Fulcher, G. (2000). The 'communicative' legacy in language testing. *System*, 28, 483-497.
- Gadd, N. (1998). Towards less humanistic English teaching. *ELT Journal*, 52(3), 223-234.
- Hamid, M. O. (2014). World Englishes in international proficiency tests. *World Englishes*, 33(2), 263-277.
- Hamid, M. O. (2016). Policies of global English tests: test-takers' perspectives on the IELTS retake policy. *Discourse: Studies in the Cultural Politics in Education*, 37(3), 472-487.
- Harris, L. R., & Brown, G. T. L. (2016). The human and social experience of assessment: Valuing the person and context. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 1-17). New York; London: Routledge.
- Hoang, N. T. H. (2018). *Test-takers' contribution to the validation of the uses of high-stakes language tests*. Unpublished PhD dissertation, the University of Queensland, Brisbane, Australia.

- Hoang, N. T. H., & Hamid, M. O. (2016). 'A fair go for all?' Australia's language-in-migration policy. *Discourse: Studies in the Cultural Politics in Education*, 38(6), 836-850.
- Hoffmann, B. (1962). *The tyranny of testing*. New York: Crowell-Collier Press.
- Kerr, P. (2007). 'Humanising'- what's in a word? *Humanising Language Teaching*, 9. Retrieved from <http://www.hltmag.co.uk/may07/mart04.htm>
- Lam, T. C. M. (1995). Fairness in performance assessment. *ERIC Digest*, ED391982. Retrieved from <http://ericae/net/db/edo/ED391982.htm>
- Luke, A. (2004). Teaching after the market: From commodity to cosmopolitan. *Teachers College Record*, 106(7), 1422-1443.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333-349.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21(3), 215-237.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 9-25). Oxford: Oxford University Press.
- O'Sullivan, B., & Green, A. (2011). Test taker characteristics In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 36-64). Cambridge UCLES/Cambridge University Press.
- Peirce, B. N., & Stein, P. (1995). Why the "Monkeys Passage" bombed: Tests, genres, and teaching. *Harvard Educational Review*, 65(1), 50-65.
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd Ed.). Cambridge: Cambridge University Press.
- Sarich, E. (2012). Accountability and external testing agencies. *Language Testing in Asia*, 2(1), 26-44.
- Shohamy, E. (2001). *The power of tests: a critical perspective on the uses of language tests*. Harlow, England; New York: Longman.
- Stoynoff, S. (2009). Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language Teaching*, 42(1), 1-40.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.

Templer, B. (2004). High-stakes tests as high fees: Notes and queries on the international English assessment market. *Journal for Critical Education Policy Studies*, 2(1), 189-226. Retrieved from <http://www.jceps.com/archives/414>

van der Heijden, J. (2013, 14 December 2013). Testing skilled migrants' English: Ridiculous and insulting. *Independent Australia*. Retrieved from <https://independentaustralia.net/australia/australia-display/testing-skilled-migrants-english-ridiculous-and-insulting,5989>

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

© Copyright rests with authors. Please cite *TESL-EJ* appropriately.