Test-takers' perspectives on a global test of English: questions of fairness, justice and validity

Check for updates

M. Obaidul Hamid[*] , Ian Hardy and Vicente Reyes

* Correspondence: m.hamid@uq.edu.au
School of Education, Social Sciences Building 24, St Lucia Campus, The University of Queensland, Brisbane 4072, Australia

Abstract

Although language test-takers have been the focus of much theoretical and empirical work in recent years, this work has been mainly concerned with their attitudes to test preparation and test-taking strategies, giving insufficient attention to their views on broader socio-political and ethical issues. This article examines test-takers' perceptions and evaluations of the fairness, justice and validity of global tests of English, with a particular focus upon the International English Language Testing System (IELTS). Based on relevant literature and theorizing into such tests, and on self-reported test experience data gathered from test-takers ($N = 430$) from 49 countries, we demonstrate how test-takers experienced fairness and justice in complex ways that problematized the purported technical excellence and validity of IELTS. Even as there was some evidence of support for the test as a fair measure of students' English capacity, the extent to which it actually reflected their language capabilities was open to question. At the same time, the participants expressed concerns about whether IELTS was a vehicle for raising revenue and for justifying immigration policies, thus raising questions about the justness of the test. The research foregrounds the importance of focusing attention upon the socio-political and ethical circumstances that currently attend large-scale, standardized English language testing.

Keywords: English language testing, IELTS, Globalization, Test-taker perspectives, Fairness, Justice, Validity

Introduction

Language test-takers have been the focus of much theoretical and empirical work in recent years. This work has been mainly concerned with their attitudes to test preparation and test-taking strategies (e.g. Cheng & DeLuca, 2011), giving insufficient attention to their views/perceptions on broader politics and ethics attending such tests. This article reports on an International English Language Testing System (IELTS) study in an Australian university, focusing particularly upon test-takers' perceptions and evaluations of the test from the perspective of the interrelated concepts of fairness, justice and validity.

Fair, just and valid use of tests and test scores are expected of high-stakes tests such as the IELTS, given that they

serve as both door-openers and gate-keepers. That is, decisions that are made on the basis of language assessments will involve allocating resources, opportunities, or rewards to some while denying these to others. (Bachman & Purpura, 2008, p. 456)

This door-opening/closing potential of high-stakes tests refers to their 'social consequences' (McNamara & Roever, 2006; McNamara & Ryan, 2011; Messick, 1989) which necessarily implicate fairness (e.g. whether test-takers are treated equally and are given equal opportunities to demonstrate their best performance), justice (e.g. how just it is to use the test for its intended (or unintended) purposes and who benefits/loses from the use of the test) and validity (e.g. how justified are the decisions that are made about test-takers based on their scores). While fairness generally refers to test-internal technical and procedural issues and validity to the logic of score-based decisions, justice refers to, among other issues, test-external policy questions. This view of justice is emphasized by Deygers (2017), Kunnan (2014) and McNamara and Ryan (2011). For example, Deygers (2017) in his recent article on the principles of justice in language testing questions:

Is it *just* for a university to demand that international L2 students meet language requirements that are not met by all L1 students, who are exempt from taking the test? Is it *just* for a country to raise the language requirements for citizenship to a literacy level that de facto excludes people who have not had access to organized education or schooling? […]. (p. 143, emphases added)

How test-takers perceive English proficiency tests that entail socio-political and ethical questions in terms of fairness, justice and validity, and what kind of experiences they have of test-taking and test impact deserve critical scrutiny, particularly given that studies have shown that test results mediate global mobility (Ahearn, 2009; Deygers, 2017; Hamid, Hoang & Kirkpatrick, 2019; Hoang & Hamid, 2017). As would be expected, the language testing literature has given significant attention to test fairness and validity; the concept of justice has also gained attention in the past decade (see Kunnan, 2014). However, while scholars and researchers engage with fairness, justice and validity in an *intellectual* sense, test-takers *experience* the consequences of different degrees of (un) fairness, (in) justice and (in) validity in very *material* ways. Therefore, understanding test-takers' perspectives/perceptions may help access their lived experiences of fairness, justice and validity which may lead to more socially responsive enactment of language testing and assessment. The growing research attention given to IELTS test-takers, but the relatively inadequate understanding of broader conceptions of fairness, justice and validity from test-takers' perspectives, were the primary motivations for undertaking the research reported in this article.

IELTS in a globalized world

IELTS is a global test of English, which is jointly owned by the British Council, IDP (International Development Program of Australian Universities and Colleges Ltd): IELTS Australia[1] and Cambridge English Language Assessment. It tests English as a

---

[1]IDP is owned by 38 Australian universities and the job website SEEK. As well as testing services, it provides international student placement services. See www.australia.idp.com.

second language (ESL) proficiency in the areas of listening, reading, writing and speaking, and reports test-takers' performances on 9-point band scores (1 = Non user and 9 = Expert user of English). Raw scores are calculated into an aggregated score for each of the four components, and then a single score, combining the results of the four modules, is assigned to each test-taker. IELTS is divided into two test types: Academic and General Training, both following the same scoring and reporting procedures. The IELTS website provides details on the testing and scoring procedures, emphasizing their fairness, reliability and trustworthiness. Since its introduction in 1989, IELTS authorities have carried out or commissioned research on various aspects of the test and research findings have led to multiple revisions in the past two decades (see Chalhoub-Deville & Turner, 2000; Stoynoff, 2009). Fairness and test impact have received considerable attention in the IELTS authorized research (see Hawkey, 2006; Hyatt, 2013).

The IELTS website claims that IELTS 'is the world's most popular … English language test' (see https://www.idp.com/global/ielts) for study, work and migration, with over 3 million tests taken in 2017. Test-takers can take the test in more than 1100 locations in over 140 countries. The test's acceptability has also been extended globally with the number of institutional test-users exceeding 9000 entities, including schools, universities, employers, immigration authorities and professional bodies in both traditional and emerging English-speaking nations. Initially introduced to assess language proficiency of ESL students seeking admission to universities in English-speaking countries, IELTS has recently been assigned a gate-keeping role for immigration and employment purposes in Australia, Canada and the UK (see Ahearn, 2009; Hamid et al., 2019; Hoang & Hamid, 2017; O'Loughlin, 2011). Educational institutions and immigration departments in these countries have set up different IELTS score requirements for prospective students and visa applicants (see Hyatt, 2013).

While the global expansion of IELTS may help enhance its technical quality and recognition, drawing upon its accumulated resources, expertise and discourses of global trust,[2] the expansion may also raise some concerns. For example, the continued expansion may affect its educational applicability, reinforcing its performative 'business-oriented' purposes (Davidson, 1993; Templer, 2004). While educational commodification is a global concern (Luke, 2004), there may be questions about whether and to what extent profit motives guiding IELTS and similar tests align with their stated goals of measuring language proficiency. There may also be concerns about whether tests driven by a profit-motive pose social, educational and ethical issues and challenges for stakeholders including test owners, researchers and language educators (see Sarich, 2012 for an examination of these issues in relation to external standardized tests in Japan).

As a transnational test operating in a globalized world, IELTS seeks standardization of the test and testing procedures. The test is produced in a center based in the UK. The hundreds of test sites located across the world have been set up as the operational units which administer the test, following written protocols for test administration produced by the center. It can be argued that this centralized operation has ensured test

---

[2]The IELTS website (https://admissiontestportal.com/en/pages/2-about-ielts/7-what-is-ielts/) notes that the 'IELTS test is trusted by over 9,000 organizations'.

fairness; no matter where test-takers are located, they receive the same test input under comparable conditions. However, while this one-size-fits-all approach may be construed as fostering 'fairness', this also ignores that test-takers come from different socioeconomic, sociolinguistic and sociocultural backgrounds with variable interests, motivations, strategies and experiences of learning and using English in different social contexts.

The test's emphasis on a particular variety of English is also reflective of its centrist tendencies. Although IELTS claims to have adopted 'international English' in response to the changing face of English (Taylor, 2006, 2010), its definition of 'international' has a narrow scope, and actually refers to what it called 'native' varieties of English; within IELTS, these are understood as British, American, Australian and New Zealand English.[3] Other varieties of English such as Indian or Malaysian or Singaporean English are yet to be substantively considered (Davies, 2009). The IELTS model of English potentially benefits those who have access to both metropolitan and 'internationally-accepted' varieties of English, and disadvantages those who speak only local varieties. Moreover, it undermines the diversity of Englishes that actually exists in an increasingly fluid world, thereby reproducing a linguistic hierarchy which discriminates against 'non-native' Englishes (Kachru, 1992).

It is not assumed that incorporating all Englishes, including 'native' and 'non-native' varieties, into the test, is problem-free, or can guarantee fairness and higher levels of performance by test-takers of different English language backgrounds (see Hamid, 2014). However, the hybridization of English and the recognition of what is called 'World Englishes' (Kachru, 1992; Kirkpatrick, 2010) have implications for global tests that have prioritized 'native' varieties of English. Concerns about language variety issues associated with IELTS raise fundamental questions about whose interests are at stake in the design of such tests:

> 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? (Brown, 2004, p. 319)

If IELTS is meant for L2 test-takers, and if the language itself has been transformed into multiple varieties globally, it may be unfair to test people in a variety that many participants are not exposed to in their social and linguistic environments. However, a more varied conception of English underpinning IELTS might be seen as inappropriate when the test purpose is taken into consideration; it could be argued that when the target language use (TLU) domain is characterized by 'native' English, this becomes an acceptable criterion upon which to judge test-takers' English capacity. In this sense, excluding 'non-native' Englishes from IELTS may appear reasonable.

Nevertheless, this latter argument seems weak because the TLU domain in 'native' English-speaking countries has now become a meeting place of students, academics and migrants from all over the world, many of whom speak a wide variety of Englishes (Batziakas, 2017). The test construct defined with reference to traditional conceptions of the TLU domain may no longer be acceptable, since the linguistic landscape has

---

[3]The website indicates: 'All standard varieties of native-speaker English, including North American, British, Australian and New Zealand English are accepted' (https://www.ielts.org/en/what-is-ielts/ielts-introduction).

significantly changed in these contexts, raising questions about whose variety of English should be included in the definition of proficiency, and how such variability may affect considerations of test performance.

Finally, the use of IELTS to control global flows of people may also be of concern. Although the danger of linguistic deficiency in border-crossing cannot be underestimated (Piller & Takahashi, 2011), the question of whether language should be given a gate-keeping role in restricting people's mobility raises ethical challenges (Deygers, 2017 as cited above; see also Capstick, 2011). If the test opens up opportunities[4]for those who are successful, it may be closing down opportunities to those who do not succeed against more standardized measures (see Bachman & Purpura, 2008).

In sum, the location of IELTS in a globalized world, its use as a gate-keeper of global mobility and the variety of English used in the test and actual Englishes used in TLU domains call for further investigation into issues of fairness, justice and validity from the perspectives of those who are directly affected by test outcomes. This article reports on data from an IELTS study to understand test-takers' perspectives on fairness, justice and validity, and to give voice to test-takers on social, political and ethical grounds more broadly.

Fairness, justice and validity in language testing

There has been a growing interest in social and political aspects of language testing associated with issues of fairness and justice and their relationship with validity. As Deygers (2017) pointed out:

> Language testers who have […] written on justice, have grappled with theoretically disentangling it from fairness, or determining its relationship with validity. (p. 147)

The interest has generally drawn on Messick's (1989) unified view of validity (e.g. McNamara & Ryan, 2011) and/or standards for educational assessment (AERA, APA,, & NCME, 2014; see also Kunnan, 2000, 2004, 2008, 2014; Taylor, 2010). However, scholars have provided varied understandings, interpretations and perspectives on the concepts of fairness and justice and how they relate to validity (see Davies, 2010; Kane, 2010; Kunnan, 2000, 2004, 2008, 2014; McNamara & Roever, 2006; McNamara & Ryan, 2011; Weir, 2005; Xi, 2010).

Kunnan is credited with highlighting the 'primacy of fairness' within 'a framework of social justice' (Kunnan, 2000, p. 1). Inspired by ethical concerns in language testing and drawing on professional standards and codes of practice, he initially defined fairness as comprising three elements: *validity*, *access* and *justice*. He subsequently revised these elements and included new items in his *test fairness framework* (Kunnan, 2004, 2008) which comprises *validity*, *absence of bias*, *access*, *administration* and *social consequences*.

For Kunnan (2004, 2008), fairness is a super-ordinate concept which subsumes validity. One implication of this conceptualization is fairness issues cannot be addressed within the framework of validity and therefore need to be investigated separately. This is the second of the three types of fairness-validity relationships discussed by Xi (2010):

---

[4]For example, the website claims that IELTS 'works around the world' and 'opens doors' (https://www.idp.com/global/ielts).

(1) fairness as an independent test quality, (2) fairness as all-encompassing, and (3) fairness as directly linked to validity. Xi's (2010) own proposal for investigating fairness is informed by the third type. While Kunnan (2010) critiques Xi's proposal for its limited understanding of fairness, Davies (2010) emphasizes the supremacy of test validation, thus rendering the linking of fairness argument to validity underpinning Xi's proposal redundant.

Kunnan's (2004, 2008) conceptualization of fairness can be seen as responsive to concerns about *procedural* fairness, but the conceptualization may also be utilized to cultivate *substantive fairness* (Kane, 2010). The former requires that test-takers be tested in essentially the same way under the same conditions, while the latter necessitates that score interpretations and test-based decisions be reasonable and equally appropriate for all test-takers.

Kunnan's (2000, 2004) encompassing notion of fairness also subsumes *justice* which he understands to be related to *societal equity* and *legal challenges*. At the same time, arguably, particularly in his earlier work, he does not seem to make sufficient distinction between fairness and justice, as rightly pointed out by McNamara and Ryan (2011). The interchangeability of the two concepts can be understood from the first of the two principles (i.e. *principle of justice* 'A test ought to be fair to all test takers') and sub-principles guiding his framework:

Sub-principle 1: A test ought to have comparable construct validity in terms of its test-score interpretation for all test-takers.
Sub-principle 2: A test ought not to be biased against any test-taker groups, in particular by assessing construct-irrelevant matters. (Kunnan, 2008, p. 14)

The principle of justice here refers to what Xi (2010) terms *comparable validity*. Kunnan here seems more concerned with what Lam (1995) calls the *equality view of fairness*. Lam (1995) also mentions the *equity view of fairness*, which is antithetical to equality because instead of being concerned with comparable procedures and outcomes, an 'equitable assessment is tailored to the individual student's instruction context and social background' (n. p.).

In short, in conceptualizing fairness, many scholars have taken what can be called a *nothing-beyond-the-test* position, even as they may believe themselves to be adopting justified approaches. Such narrower approaches provide little scope for asking questions about the socio-political purposes of tests.

In contrast, McNamara and Ryan (2011) have taken a *beyond-the-test* position and have used *justice* in a specific sense based on Messick's (1989) model of test validity which is to be distinguished from fairness. As these authors explain:

By *fairness*, here we mean the extent to which the test quality, especially its psychometric quality, ensures procedural equality for individual and subgroups of test-takers and the adequacy of the test representation of the construct in test materials and procedures. (McNamara & Ryan, 2011, p. 163)

McNamara and Ryan (2011) argue that what is missing from this narrower definition of fairness is a basic test-external issue that refers to test consequences as well as social

values that tests embody. The term *justice* is reserved for this particular aspect of fairness; on such a rendering:

> any problem with the test may not inhere in its quality, but its very existence and use in the first place, no matter how technically sophisticated and 'fair' in the narrow sense it may be. (McNamara & Ryan, 2011, p. 164)

Drawing upon McNamara and Ryan (2011) and Sen (2010), Deygers (2017) identified two conceptual distinctions between fairness and justice. First, justice depends on fairness, but fairness cannot be a proxy for justice. Second, the scope of justice should go beyond test-takers and encompass the impact of tests in society because the potential inequities brought about by tests might not have existed before their introduction. Kunnan's (2014) latest exposition of fairness and justice that draws on Rawls (2001) and Sen (2010) outlines similar test internal-external distinctions which appear to be aligned with McNamara and Ryan's (2011) conceptualizations. As he explains, this work:

> attempts to provide principled bases for fairness and justice as applied to the institution of assessment. It does this by applying the *idea of fairness as relating to persons—how assessments ought to be fair to test takers—and the idea of justice as relating to institutions—how institutions ought to be just to test takers*. (Kunnan, 2014, p. 2, italics original)

McNamara and Ryan acknowledge that the social concerns that are brought within this scope of *justice* are derived from their interpretations of Messick's (1989) facets of validity. Messick defined validity as 'an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment' (p. 13; emphasis original). This view of validity marks a departure from earlier conceptualizations which considered validity as a property of tests as indicated by the concept of 'construct validity' that refers to whether tests measure what they intend to measure. For McNamara and Ryan (2011), construct validity is an element of fairness as it refers to technical issues. To them, neither fairness (as discussed above) nor validity can guarantee justice. Justice demands validity which, in turn, is contingent on fairness, but a fair and valid test may also be imposed unjustly on test-takers (Deygers, 2017).

McNamara and Ryan's view of justice owes much to Shohamy's (2001) work on critical language testing (CLT) which aims to expose the 'potential and real injustice of tests, rather than of critiquing their psychometric qualities as the principal source of their illegitimacy' (McNamara & Ryan, 2011, p. 165). Their view is also comparable to the *principle of beneficence* guiding Kunnan's (2008) later test fairness framework, which states that a 'test ought to bring about good in society, that is, it should not be harmful or detrimental to society' (p. 14). The *test context framework* that Kunnan (2008) introduced to complement his earlier *test fairness framework* incorporates this principle which is 'necessary to examine tests and testing practice from a wide context in order to more fully determine

whether and how these tests are beneficial or detrimental to society' (p. 14). Kunnan's (2014) revised *principle of justice* also emphasizes that an assessment institution 'ought to be just and bring about benefits in society' (p. 8).

Based on the discussion above and drawing mainly upon McNamara and Ryan (2011), we define the three concepts in the following way:

Fairness, as referring to test-internal technical quality which ensures that test-takers are treated equally and are given equal opportunity to demonstrate their best performance.
Validity, as referring to the adequacy, appropriateness and justification of decisions made on the basis of test scores.
Justice, as referring to test-external issues to ensure that the use of tests for their stated purposes is justified and their introduction does not have harmful impacts on society.

Fairness, justice and validity are interdependent qualities which are not adequate on their own; affecting one necessarily affects the other two. Achieving an optimum balance between them would be a reasonable goal keeping in mind the usefulness and practicality of testing.

Fairness, justice and validity from test-takers' perspectives

The importance of understanding test-takers' perspectives has been emphasized in the literature. For example, Weir's (2005) socio-cognitive framework for an evidence-based validation model has an important focus on how test-takers' physical, psychological and experiential characteristics can be taken into consideration. O'Sullivan and Green (2011) provide further details on these domains of test-taker characteristics. However, these models, as well as much of the empirical work, are guided by fairness and validation in a narrow sense. Peirce and Stein (1995) demonstrated how a multiple-choice test conditioned a group of Black students in a South African school to submit to meanings expected by the test regime at the expense of their own meanings, histories and experiences. However, guided by validity in a procedural sense in the selection of test content, they gave little attention to the dehumanizing potential of tests more broadly. Working with IELTS test-takers and teachers, Hawkey (2006, 2008) investigated how these stakeholders perceived the test in terms of, among other issues, fairness, difficulty and test anxiety. Of note was the inclusion in the survey of this structured question: *Do you think IELTS is a fair way to test your English proficiency?* While the study's engagement with students and teachers from a fairness point of view is welcome, again, it can be argued that the focus here was on procedural fairness, although the way the question was posed left room for multiple interpretations.

Compared to such test-internal fairness issues, Ahearn's (2009) case study of an IELTS test-taker from Korea problematizes the use of the test for academic, political (immigration) and commercial purposes (see also Hoang & Hamid, 2017). Similarly, Chik and Besser's (2011) case study involving young test-takers, their parents and school principals highlights private uses of commercial English language tests and their

educational and social consequences. More importantly, the researchers argued that the actual beneficiaries of the tests were testing agencies:

> However, the real power is held not by the stakeholders but by the testing agencies. These agencies are profiting from the parental worry over their children's future educational opportunities. (Chik & Besser, 2011, p. 88)

This article reports on data from an IELTS study to add to this body of work on socio-political and ethical issues in language testing from test-takers' perspectives.

Context, participants and data collection methods

The article is based on a larger study into language test-takers' test-taking experiences, undertaken at a major, metropolitan Australian university[5]where the authors work. The participants (N = 430, female = 45.3%, male = 54.7%) were from 49 countries including five 'native' speakers of English (three from UK and two from Ireland). About 60% of the test-takers were studying at the time of the research and 70% took the test for study purposes. Well over half of the sample took the test at least twice and a considerable proportion of participants took the test three, four or five times. Finally, the largest proportion of test-takers (about 80%) was from the more successful group who scored between 6.5 and 9 on the IELTS band scale.

The main instrument for data collection was a survey for the larger study available in both online and paper versions (see Hamid, 2014 for details). The survey questionnaire was organized into three sections in which the first asked for information on test-takers' national backgrounds, current occupational/academic status, the reasons for taking IELTS and scores obtained. The second part contained 40 structured items on test-taking experience, fairness and validity and the socio-politics of IELTS, and the varieties of English used in the test. The final section contained an open-ended question inviting participants to comment on other issues and/or provide suggestions for test improvement.[6] The present article is based on test-takers' responses to this open question. These responses were particularly important given that test-takers freely expressed their views on many aspects of the test without being constrained by space or response format.

Of the 430 survey participants, 343 (80%) volunteered written comments, producing a corpus of 18,500 words. On average, each response contained 53 words, the longest one comprised 536 words, and the shortest one just two words. In part, and following procedures for qualitative coding and content analysis (Corbin & Strauss, 2008; Dörnyei, 2007), the responses were read repeatedly and coded at the level of phrases, sentences and paragraphs, following a broadly inductive approach. A colleague, whose research focuses on test-takers' perspectives on international proficiency tests including

---

[5]Ethical clearance for the research was obtained from this and another university in the same city.
[6]The following prompt was used for the open-ended comments:

> 'Do you have anything else to say about the test? Would you suggest any changes to the test? Please make your response as long as you like.'

IELTS and TOEFL, read all responses and the codes and she agreed with over 98% of the coding. At the same time, and reflecting how research is also always an active process reflecting the interests of the researchers, and not some sort of 'objective' exercise, key codes were analysed in light of relevant theorizing and literature on IELTS, pertaining to conceptions of fairness, justice and validity, as outlined above. In this way, the data analysis process involved simultaneous processes of engagement with theorizing and data involving both inductive and deductive processes, in light of our particular focus.

More broadly, the research is underpinned by critical and constructivist views that recognize people's voice and agency in making sense of their experiences, and of the necessarily socially situated and embedded nature of such responses. Positivistic views may flag subjectivity and potential bias in the representations of selves and personal experiences; these 'perception' data may also not be considered hard evidence from such epistemological perspectives. However, being guided by perspectives that respect individual agency, and research as an always active process on the part of the researcher (Bourdieu & Wacquant, 1992), we believe that 'people's reasons and accounts provide evidence whose status is ontologically real' (Corson, 1998, p. 63, referring to Bhaskar, 1986), even as this 'reality' is also simultaneously an act of construction through the research process. Participants' voices and representations or their 'perceptions' are important resources to help develop an understanding of fairness, justice and validity from test-takers' lived experiences.

Results/findings

Test-takers' perceptions covered a variety of issues related to evaluations of the test (e.g. 'I think it is ok and I have nothing to complain about it'), their experience of taking the test ('Taking IELTS was not a pleasant experience'), the test fee, the duration of its score validity, the relevance of 'non-native' varieties of English, their perceptions of the test's ability to measure language proficiency, their suggestions for test improvement and issues of fairness, justice and validity. In this article, their responses to these latter issues are of particular interest, and were expressed in relation to whether IELTS was construed as beneficial, whether IELTS was seen as an accurate measure of English proficiency and whether decisions made based on test-scores were justified and the purpose of the test. We endeavour to show how notions of fairness, justice and validity were intertwined within these themes, even as we employ a degree of analytical separation to assist the reader in following our arguments.

IELTS as a beneficial test?

A generally positive evaluation of the test was evident in 51 comments[7] by respondents, compared to 13 comments which were negative. Moreover, about half a dozen test-takers underscored the necessity of IELTS for its stated purposes. This can be seen in the extract from R115[8] who wrote:

---

[7]That is, manual coding of the data found 51 references to positive evaluations of the test.
[8]Respondent codes such as these are used throughout to ensure anonymity.

It is therefore imperative to have good English (good IELTS score) before even starting their [international students'] studies. A[s] it currently stands, I've seen a lot of students graduating with abysmal level of English [...] For this reason, I support the immigration policy that requires you to retake IELTS when applying for skilled migration visa, *even if you graduated in Australia*. (emphasis added)

From this respondent's perspective, the test was beneficial and justified (just) because it provided an 'accurate' measure of test-takers' language ability (good IELTS scores equal good English). This view seems to represent that fairness and validity equal justice, which McNamara and Ryan (2011) as well as Deygers (2017) might consider a misrepresentation. Three other test-takers not only believed in the test's ability to measure their language proficiency but also argued that IELTS and TOEFL (Test of English as a Foreign Language, another global test of English) should be merged so that there is one English test in the world, with universal standards and assessment criteria. On such renderings, broader conceptions of beneficence guiding Kunnan's (2008) test fairness framework, with its emphasis upon how a 'test ought to bring about good in society', and that 'it should not be harmful or detrimental to society' (p. 14) did not bear upon these respondents' understandings of the test (see also Kunnan, 2014).

The necessity of maintaining such standards—understood as a requirement for 'fairness'—was the basis for the rejection of 'non-native' Englishes, and these respondents' insistence on British English. Eighteen respondents were against alternatives to standard/ 'native' English varieties, compared to eight respondents who supported their inclusion. Only one respondent upheld a mixed position that suggested that 'non-native' varieties of Englishes could be included in speaking, but not in the other components (see also Hamid, 2014). The dominance of 'native' English language varieties, but alongside a considerable advocacy for other Englishes, reflects a point of possibility for fostering a more responsive approach to the plurality of Englishes that actually characterizes English speech and writing throughout the world (Kachru, 1992; Kirkpatrick, 2010). Perhaps a broader conception of validity was evident in such a response, associated with notions of adequacy or appropriateness of inferences pertaining to fairness in relation to test scores, as evident in McNamara and Ryan's (2011) and Messick (1989) understandings.

IELTS as an accurate measure of English proficiency?

Nevertheless, and at the same time, a large proportion of respondents perceived that the test did not provide an accurate measure of their language abilities. There were 50 respondents who asserted that their language ability was not reflected in their test scores, thus problematizing the assumption that good English = good scores. Such an assertion raised questions about fairness as well as validity (McNamara & Ryan, 2011). Scepticism of varying degrees was expressed in different ways. For example, R11 evaluated the test positively, but still did not believe that it measured what it purported to measure:

I personally believe that the structure and design of present IELTS examination is good. But I think that it does not measure one's ability in English. (R11)

This juxtaposition of a positive evaluation in general, followed by scepticism towards the test's technical ability, can be subjected to various interpretations. It is plausible to

see the positive assessment as part of an evaluation discourse that starts with vague positive comments before pointing to specific concerns. This can be seen more clearly in the perspective of R69, who not only took a pro-IELTS position but also hid his/her scepticism behind more optimistic initial suggestions:

> I do not want to go against the IELTS test. But, I want to say that the test modules should be designed [in] such a way that it really helps to check the test takers' English proficiency. It means that if a person scores a band score, it should reflect his [sic] real skill of English. Thirdly, the test should be designed in such a way that persons scoring [the] same band score should have the same level of proficiency. (R69)

R69 made three suggestions: (1) IELTS should test test-takers' English proficiency (construct validity); (2) there should be a meaningful correspondence between IELTS bands and ability (fairness in the sense of psychometric quality); and (3) IELTS should guarantee *comparative validity* for all test-takers. Concerns related to fairness and validity and, by implication, justice underlying each of these suggestions were voiced by other participants as well. There was a perception that IELTS was a test of 'test-wiseness', rather than of English ability. R222, a British test-taker, who took the test in 2011 for immigration purposes, observed:

> You need to know exactly how the test works to score highly, which means it isn't testing your English ability—it's testing your ability to pass IELTS.

R324, an ESL speaker, expressed similar concerns: 'having a high IELTS score doesn't mean the test-taker's [English] level is high; maybe his [sic] English is not that good but only knew how to do the test'. The value of the test was critiqued here, although in terms of test-internal technical qualities (McNamara & Ryan, 2011). Test-wiseness is also implied by R293, who questioned the test's use by suggesting that despite having passed through the IELTS hurdle, many people struggle to communicate in the host society:

> The test has recently lost its value (or both reliability and validity) as it has claimed. It doesn't really reflect how proficient a test taker is [...] A score doesn't mean a real score … it is a minimum license for one to get into an English-speaking environment. Is what happens in a test similarly what happens outside? I'm doubtful with its authenticity. (R293)

However, there was also a perception that the test did not support test-wiseness. For example, R188, an IELTS instructor, recounted:

> I am not sure if the result of the IELTS test is an accurate assessment of a person's English proficiency [...] Some of my students did not achieve the score they needed at their first attempt and they decided to take the test the second time in the hope of improving their score. However, after a lot of days taking practice tests and revising for the exam, the results of their second IELTS test were even lower than the first exam.

Issues of reliability that implicate fairness and validity raised by this respondent were also mentioned by other respondents. For example:

> I would say that IELTS exam is not a very good criterion to test peoples' language abilities. As you know many people take the exam twice in a row just in two weeks and their first score has a significant difference with the second one. In my opinion it shows that the test is not reliable. (R350)

The test re-take policy[9] (see Hamid 2016) also provided a context for participants to question the reliability and validity of the test. Participants perceived differences in test scores across two sittings as erratic, and beyond comprehension:

> For example, I took the IELTS test three times in the last two years looking for a 7 in all the bands for immigration purposes. The first time I got 8 in writing and 6.5 in speaking. The second time (around 6 months later) my results swapped: I got 8 in speaking and 6.5 in writing. The third time, I finally got 7 in both components. Does it make any sense? It doesn't for me... (R271)

As the participants reported, their scores changed substantially across the four components to such an extent that they did not know where their difficulties lay, forcing them to keep repeating the test until they obtained the desired scores in all components in the same sitting. One test-taker, who repeated the test 18 times in 2 years, noted:

> [...] my overall score was ok but I got slightly less score in a particular module. But not consistently in one module. For example one time I got 6.5 in writing but 5.5 in speaking but in the later test, my score is OK in speaking but I got less in writing. So this is a problem. (R211)

This respondent did score band 7, his target, in all four components, but he did not score all 7 s in the same sitting. Such situations, which were common in the data, illustrate how technical issues (e.g. reliability and fairness) could be linked to justice issues (e.g. use of tests) through the activation of validity issues (e.g. inappropriate or inadequate score-based decisions).

IELTS for economic and political purposes?

Several participants perceived the test as valid, even as others expressed reservations, including how its educative functions appeared to be dominated by more economic and regulatory political imperatives. While some test-takers believed that the test opened doors to opportunities as evidenced by research which showed how IELTS

---

[9]The re-take policy is described as follows in the Information for Candidates document on the IELTS website: 'There are no restrictions on re-taking IELTS. If you do not get the result you wanted, you can register for another test as soon as you feel you are ready to do so'. When re-taking IELTS, candidates have to redo all four components, including the ones in which they had satisfactory results from a previous sitting.

scores had an impact on people's life and career choices (Ahearn, 2009; Hoang & Hamid, 2017), candidates' responses also indicated that test owners had more opportunities created by the test than test-takers. By opportunities, test-takers referred to test owners' commercial gains (cf. Chik & Besser, 2011). This was clearly articulated by R362 who considered that IELTS was a money-making venture:

> Whichever area the test taker had a lower band score should only be the area to re-take. Because the test is so costly, I perceived it to be income generation by the organisation and making it hard and tough for skilled professionals who have higher qualifications than the lazy English speaking natives here in Australia.

This participant pointed to an issue that raised questions of fairness and particularly justice in relation to the test's business-oriented, money-making motives. This commercial use of the test was emphasized in more than a dozen responses, including:

- 'a money-producing machine for the native countries, i.e. the UK and Australia' (R2),
- 'the IELTS test is one of the main way [s] to make money for the test producer' (R138)[10]
- 'Although the test aims to assess English proficiency, the main objective is [to] make money' (R163)
- 'This test should not be used for the purpose of making money and profit'. (R190)

While the material consequences of linguistic deficiency in migrant societies should not be underestimated (Piller & Takahashi, 2011), the focus of the respondents here was largely economic. They pointed out that the commercial motives explained some of the policy justifications behind IELTS, including in relation to immigration issues. An Irish test-taker, who spoke English as his mother tongue, but was still required to take the test for the point-based immigration system (see Endnote 13), explained:

> Despite this, I was forced to take the IELTS test as part of my application for a skilled migration visa to Australia. I think this is madness, and a waste of money. I do think that in this case it is purely a money making exercise and nothing else. (R302)

Related to this, a few participants perceived that the test was purposefully made difficult so that test-takers had to repeat it, and the high costs would add to the test-owners' income. van der Heijden (2013) provides an estimate of how 150 IELTS test-takers in one test center in Australia contributed over AUD$50,000 to revenue in a single day. That IELTS is deliberately made difficult may not be an acceptable line of argument, but it was evident in the data. As another respondent noted:

---

[10]Test producers, test owners and test users (e.g. Australian universities) are often conflated in the test-takers' responses. Although the IELTS website makes the joint ownership of the test clear, because these owners are British and Australian institutions, often test-takers referred to these two countries, in place of these specific institutions.

> After doing the test a couple of times I have just realised what a biggest scam the Ielts [sic] is. They are making it extremely hard to pass so that they can rip off money from the students. (R208)

Similarly, the 2-year validity (or *shelf life* of test scores),[11] which appeared unreasonable to many respondents, was equally seen as related to business motives. R361 explained that staying in an English-speaking country should help people improve their English. This view seems to be supported by evidence.[12] It is noted that if they entered the country the first time 'proving' their English proficiency, it did not make sense to take the test again, for academic or immigration purposes. Unjustified use of the test involving use of the test to apply for residency (McNamara and Ryan) was also perceived as problematic by other respondents:

> For example, some international students who already prove their language skills to their educational institutions still have to take it because it's a requirement of the dept [Department] of immigration. (R401)

This respondent perceived the initial use of IELTS for (in this case) academic purposes as appropriate, but it was seen as inappropriate to have to re-sit the test for student visa processing purposes (see Capstick (2011) for further insights on this issue). In other words, R401 pointed to an unjust use of IELTS as perceived by test-takers. This viewpoint was also shared by another respondent who, while stating that IELTS was 'the best possible English test', felt the test was a burden to 'non-native' speakers of English in the way it was currently being used:

> It does have minor flaws, however if anyone asks me to name the best possible English test, I would not hesitate to recommend IELTS. Unfortunately, IELTS is currently being abused by the system/government/immigration to put too much burden on non-native speakers, such as setting the bar a little too high for them. Asking overall and/or each band score(s) of 7.0 seems to be unreasonable. (R398)

R398 expressed his/her faith in the quality of IELTS and its fairness in the sense of equality, but he/she considered it to be unfair—and therefore unjust—as a social and

---

[11]The relevant policy is mentioned in the Information for Candidates on the website, although no reason is provided for 2 years. It is also unclear whether this policy is needed for the IELTS partners or the test users (see Hamid et al., 2019).
[12]For example, the Australian Council of TESOL Associations in its formal submission to the Senate Legal and Constitutional Affairs Committee regarding the Australian Citizenship Legislation Amendment Bill 2017 noted:

> there is no evidence that the English language skills of permanent residents decline over time. There is credible evidence to the contrary, not least from the 2011 ASRG [Australian Survey Research Group] report, which was commissioned by the then Immigration Department. (http://www.tesol.org.au/files/files/577_Sub292_-_ACTA_sub_to_Citizenship_Inquiry_July_2017.pdf)

However, it has also been reported that some international students do not improve English proficiency even after their graduation from Australian universities (see Burton-Bradley, 2018).

political tool for controlling the flow of international students as potential immigrants into Australia. It seemed to reflect the door-opening/closing potential of high-stakes tests as part of the 'social consequences' of such tests (McNamara & Roever, 2006; McNamara & Ryan, 2011; Messick, 1989) and which necessarily implicate notions of justice, and beyond more technicist notions of fairness of the test. This gate-keeping function was perceived as problematic by another respondent who raised the issue of fairness and justice differently by referring to 'native' speakers of English and their own lack of English proficiency:

> IELTS is quite a paradox. Certain governments have used it as a requirement for migration and educational purposes. Yet, if they were to apply the same test on their own citizens, their own citizens would not be able to pass these tests. (R311)

This view reflects Deygers's (2017) previously cited question about the justness of the demand of the levels of English proficiency from L2 speakers which may not be met by some L1 speakers of English. R380 provided validation of this point:

> I know I am international student but [...] there are a lot of locals who couldn't even spell properly, how come they are expecting us to have the perfect English wherein [sic] some of them cannot even spell properly. Will they pass the IELTS?

By requiring 'native' speakers to take IELTS for immigration purposes,[13] some sort of fairness seems to have been established (even as this was contested by 'native' English speakers, as indicated earlier). However, test-takers, who were yet to be convinced of many aspects of IELTS policy and its use for overly commercial motives, rallied against what they perceived as the injustice of the IELTS process by pointing to a community of reference within the TLU society whom they believed to be less proficient in English than themselves. Test-takers' insights strongly indicated that their level of proficiency should be more fully taken into consideration in defining the IELTS construct, establishing its purposes and evaluating its performance standards.

Discussion and conclusions

Investigating how Global English tests such as IELTS impact test-takers' lives and global mobility and how test-takers perceive the processes and outcomes of test design, administration and use are more than an academic exercise; it is an imperative for social justice, not only between groups of test-takers but also between test-takers and test-owners/test-users who are locked in a relationship of inequality (Deygers, 2017). While the commercial viability of tests, the pragmatics of testing and the limits of testability including test qualities (Bachman & Palmer, 1996) need to be appreciated, the science and technology of tests and professional standards alone may not be adequate for guiding test design, administration and use (Hamid, 2016). Critical reflection on the operation and the intent of IELTS points to the complexity of fairness, justice and validity, as exemplified in the present study. To date, test-takers' perceptions of IELTS, in relation to a broader, more encompassing conceptions of fairness, justice and validity

---

[13]The new points system for skilled migration to Australia provides points for English proficiency for both 'native' and 'non-native' speakers of English.

have not been given adequate attention in the literature. This article provides an example of engagement with test-takers in the hope that test-takers' experiences and perspectives will be more fully taken into consideration in test design, administration and use (Dimova, 2012; Green & Andrade, 2010).

On a narrower rendering of issues of fairness, the test-takers were critical of IELTS even as a large proportion initially indicated, from their perspective, that the test was 'fair' in the way in which it sought for all participants to take exactly the same test. This 'sameness' refers to procedural fairness which was seen as ensuring a level of equivalence that could not otherwise be achieved. However, and, at the same time, an overwhelming proportion expressed the view that the test did not provide an accurate measure of their proficiency, raising questions about the reliability of scores and about fairness and validity as a consequence. Several explanations emerged from the data for the perception that the test did not test participants' actual language proficiency. First, personal experiences of test repetition showed substantial variations in scores across test sittings. Secondly, test-takers' experiences of engaging in English in the host society enabled them to understand that higher scores did not necessarily mean higher levels of performance in language use contexts and vice versa. Thirdly, they believed that the test did not guarantee comparative validity, meaning that two test-takers with the same scores did not necessarily demonstrate the same level of English proficiency. Such views may reveal the 'potential and real injustice of tests' (McNamara & Ryan, 2011, p. 165), and how test-takers felt that they had to navigate these vicissitudes, even as they appeared to make little sense to them. Test-takers' self-reported experiences reflected the gap between testing in theory (as reflected in scores), and their actual proficiency, and real-world language use.

Most significantly, participants' (subjective) responses reveal how a broader conception of justice is restricted by the testing practices and processes associated with IELTS. While some participants argued that a sense of procedural fairness was evident in the way the test was constructed, and how it was enacted, others were critical of what they construed as a set of practices designed for external purposes—namely to generate profits from an international student market, and to serve as a potential barrier to access, and restriction upon immigration. Notions of beneficence guiding Kunnan's (2008) test fairness framework, with its emphasis upon how a 'test ought to bring about good in society', and that 'it should not be harmful or detrimental to society' (p. 14), were sorely tested by an evaluation framework that seemed to be more driven by extraneous motives of profit, as perceived by the test-takers, than by efforts to foster the sorts of cosmopolitan, harmonious and diverse societies that could be cultivated through providing opportunities for respectful engagement with individuals and groups from rich and varied cultural backgrounds. Test-takers reported that they experienced a lack of fairness and a sense of injustice that seemed to problematize the technical excellence supposedly associated with the 'validity' of IELTS.

Reflecting McNamara and Ryan's (2011) conception of justice, test-takers' responses indicated a broader conceptualization of justice than simply a technicist focus upon issues of equality as 'sameness' in relation to testing alone. Such responses also reflect how processes of recognizing difference in robust, activist ways (Fraser, 1997) was severely lacking, and that the tests seem to have been 'stacked against' them, even as participants simultaneously perceived that the test played a crucial role in their pursuit

of opportunities in the countries in which they studied, and to which many may have wished to move upon completion of their studies. Participants' responses also revealed that ultimately it was test-owners who were the real beneficiaries of the testing process. As Chik and Besser (2011) indicated with reference to similar international tests for young English learners, profit maximization was considered to be behind the test-retake policy of IELTS which was also, as some test-takers believed (or perhaps misbelieved), deliberately made difficult for the purpose of maximizing profits. The psychological, social and economic costs that they reported they incurred in taking IELTS also indicated validity concerns, particularly when participants expressed reservations about 'native' English speakers' capacity to succeed in the test. Participants considered it unjust to impose a level of proficiency on ESL speakers which was less than evident within the TLU community more broadly. Test-takers may have blurred the boundaries of the domains of testers and test-users, but it may be testers' responsibility to develop test literacy and awareness among test-takers. Testing agencies may also need to improve test-takers' limited understanding of the interdependence of fairness, validity and justice—specifically, to make them understand how achieving a reasonable balance of the three concepts may require a principled compromise of each.

While more research involving larger samples of test-takers using varied methods of data collection is needed beyond participants' perceptions, these powerful insights (however subjective) deserve greater attention, and may help inform actions on the part of IELTS providers, governments, and testing researchers. For IELTS authorities in particular, the research reveals a need to consider and explain and perhaps reconsider (1) the rationale behind the test, including the period of validity; (2) the allowable error margins; (3) technicist notions of comparative validity; and (4) how to enhance assessment literacy among test-takers. It is also necessary to rethink the re-take policy and establish ways of communicating with test-takers to address perceived concerns about IELTS and its broader goals.

The research revealed the complexity and problematics that attend standardized English test-taking processes such as those associated with IELTS, and how such tests can be seen as part of broader processes of homogenizing, global English testing practices. It reinforces the need to sustain critical engagement with the nature of such practices more generally. The ways in which test-takers perceived and experienced the fairness and validity of the tests, and the extent to which these experiences reinforced and challenged notions of justice, revealed such testing practices as requiring much closer scrutiny, particularly in relation to both motives and processes that drive their use in a globalizing world.

Authors' information
M. Obaidul Hamid (Orcid: 0000-0003-3205-6124) is Senior Lecturer in the School of Education, The University of Queensland. Dr. Hamid's research focuses on the policy and practice of TESOL education in Asia. He is co-editor of *Language planning for medium of instruction in Asia* (Routledge; 2014).
Ian Hardy (Orcid: 0000-0002-8124-8766) is Associate Professor in the School of Education, The University of Queensland. His research focuses on the politics of educational policy and practice, with particular attention to the nature of teachers' work and learning. He is author of *The politics of teacher professional development*: *Policy, research and practice* (Routledge; 2012).
Vicente Reyes (Orcid: 0000-0002-1539-1839) is Senior Lecturer in the School of Education, The University of Queensland. His research focuses on educational transformations, technology innovations in education and comparative education reform. His latest book is entitled *Mapping the Terrain of Education Reform*: *Global Trends and Local Responses in the Philippines* (Routledge; 2016).

References
AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association and National Council on Measurement in Education.
Ahearn, S. (2009). 'Like cars or breakfast cereal': IELTS and the trade in education and immigration. *TESOL in Context, 19*(1), 39–51.
Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
Bachman, L. F., & Purpura, J. E. (2008). Language assessments: Gate-keepers or door-openers? In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 456–468). Malden: Blackwell Publishing.
Batziakas, B. (2017). Communicative practices in English as a lingua franca interactions: Some examples from Asian university students in London. *Asian Englishes, 19*(1), 44–56.
Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.
Bourdieu, P., & Wacquant, L. (1992). *An invitation to reflexive sociology*. Chicago: The University of Chicago 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.
Burton-Bradley, R. (2018). *Poor English, few jobs: Are Australian universities using international students as 'cash cows'?* Australian Broadcasting Corporation Retrieved from https://www.abc.net.au/news/2018-11-25/poor-english-no-jobs-little-support-international-students/10513590.
Capstick, T. (2011). Language and migration: The social and economic benefits of learning English in Pakistan. In H. Coleman (Ed.), *Dreams and realities: Developing countries and the English language* (pp. 1–23). London: British Council.
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.
Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment, 16*(2), 104–122.
Chik, A., & Besser, S. (2011). International language test taking among young learners: A Hong Kong case study. *Language Assessment Quarterly, 8*, 73–91.
Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks: Sage.
Corson, D. (1998). *Language policy in schools: A resource book for teachers and administrators*. Florence: Tylor and Francis.
Davidson, F. (1993). Testing English across cultures: Summary and comments. World Englishes, 12(1), 113–125.
Davies, A. (2009). Assessing world Englishes. *Annual Review of Applied Linguistics, 29*, 80–89.
Davies, A. (2010). Test fairness: A response. *Language Testing, 27*(2), 171–176.
Deygers, B. (2017). Juts testing: Applying theories of justice to high-stakes language tests. *ITL-International Journal of Applied Linguistics, 168*(2), 143–163.
Dimova, S. (2012). Matura's rocky road to success: Coping with test validity issues. In I. Csépes & D. Tsagari (Eds.), *Collaboration in language testing and assessment* (pp. 143–157). Frankfurt: Peter Lang Verlag.
Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
Fraser, N. (1997). *Justice interruptus: Critical reflections on the 'postsocialist' condition*. New York and London: Routledge.

Green, B. A., & Andrade, M. S. (2010). Guiding principles for language assessment reform: A model for collaboration. *Journal of English for Academic Purposes, 9*, 322–334.

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: Testtakers' perspectives on the IELTS retake policy. *Discourse, 37*(3), 472–487.

Hamid, M. O., Hoang, N. T. H., & Kirkpatrick, A. (2019). Language tests, linguistic gatekeeping and global mobility. *Current Issues in Language Planning, 20*(3), 226–244.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS Progetto Lingue 2000*. Cambridge: UCLES/Cambridge University Press.

Hawkey, R. (2008). An impact study of a high-stakes test (IELTS): Lessons for test validation and linguistic diversity. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity* (pp. 215–228). Cambridge: UCLES/Cambridge University Press.

Hoang, N. T. H., & Hamid, M. O. (2017). 'A fair go for all?' Australia's language-in-migration policy. *Discourse, 38*(6), 836–850.

Hyatt, D. (2013). Stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK. *Journal of Further and Higher Education, 37*(6), 844–863.

Kachru, B. B. (1992). World Englishes: approaches, issues and resources. *Language Teaching, 25*(1), 1–14.

Kane, M. (2010). Validity and fairness. *Language Testing, 27*(2), 177–182.

Kirkpatrick, A. (Ed.). (2010). *The Routledge handbook of world Englishes*. New York: Routledge.

Kunnan, A. (2000). Fairness and justice for all. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge: UCLES/Cambridge University Press.

Kunnan, A. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.), *European year of languages conference papers, Barcelona* (pp. 27–48). Cambridge: Cambridge University Press.

Kunnan, A. (2008). Towards a model of test evaluation: Using the test fairness and test context frameworks. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity* (pp. 229–251). Cambridge: Cambridge University Press.

Kunnan, A. (2010). Test fairness and Toulmin's argument structure. *Language Testing, 27*(2), 183–189.

Kunnan, A. (2014). Fairness and justice in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1–17). New York: Wiley.

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., & Roever, C. (2006). *Language testing: The social dimension*. Malden: Blackwell.

McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian citizenship test. *Language Assessment Quarterly, 8*, 161–178.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan.

O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly, 8*, 146–160.

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.

Piller, I., & Takahashi, K. (2011). Language, migration and human rights. In R. Wodak, B. Johnstone, & P. Kerswill (Eds.), *The sage handbook of sociolinguistics* (pp. 583–597). London: Sage.

Rawls, J. (2001). *Justice as fairness: A restatement* (E. Kelly, Ed.). Cambridge, MA: Harvard University Press.

Sarich, E. (2012). Accountability and external testing agencies. *Language Testing in Asia, 2*(1), 26–44.

Sen, A. (2010). *The idea of justice*. London: Penguin.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, 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. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal, 60*(1), 51–60.

Taylor, L. (2010). Setting language standard for teaching and assessment: A matter of principle, politics, or prejudice? In L. Taylor & C. J. Weir (Eds.), *Language testing matters: Investigating the wider social and educational impact of assessment* (pp. 139–157). Cambridge: UCLES/Cambridge University Press.

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). Available at: http://www.jceps.com/archives/414.

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

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

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147–170.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.