Trust the test: Score-user perspectives on the roles of language tests inprofessional registration and skilled migration

Susy Macqueen, Australian National University
John Pill, Lancaster University
Ute Knoch, University of Melbourne

Abstract

English language proficiency is a deciding factor in the life opportunities of many thousands of applicants for Australian skilled migration every year. This paper focuses on the perspectives of professional bodies that use English language tests in their decisions. Taking an interpretative approach, we explore the meanings that policy makers from these organisations ascribe (as score users) to test standards (cut-scores) so that we can better understand the uses of test scores in migration policy. The policy narratives we observed around the use of test scores describe the need to manage large numbers of applicants, to assure a level of English proficiency for high-risk professional communications, to provide an objective assessment that is separate from any assessment of professional competence and to maintain consistency of standards with other bodies. These views are contextualised with other relevant information, particularly that available from test providers, who are key players in the test-using interpretive community. We observe that particular tests and their standards become trusted and entrenched in policy, using the apparently simple semiotics of scores. Concomitantly, trust in tests is nurtured by test marketing. These tendencies warrant attention from test researchers, providers and score users.

Key words: professional registration, skilled migration, test mandate, test standard, risk, trust

Introduction: Meanings of test scores

Test scores are both technical and social phenomena. At one level, score meaning is generated through design and technical properties, and at another, score meanings are generated socially (McNamara, 2012). This paper examines the social lives of test scores in which scores offer solutions to perceived policy needs and simple ways of understanding and communicating about language proficiency. The rationale for such an exploration is that discourse about tests is itself consequential in how and why tests (and particular test scores) are used as policy tools. We take an interpretative approach to examine test use in a sociohistorical context commencing with the rise of commercial English language testing in Australian skilled migration.

Context

With the growth in numbers of students applying to study outside their home countries, English language tests were developed specifically to assess the readiness of international students to attend English-medium universities. Two well-established English language tests developed for this specific purpose are the International English Language Testing System (IELTS), used particularly in the UK and Australia, and the Test of English as a Foreign Language (TOEFL), used particularly in the USA. A precursor to IELTS was first administered in 1980 and the present version (with some amendments) has been in use since 1995. The TOEFL was first administered in 1964 and the predominant version in current use, the Internet-based Test (iBT), was introduced in 2005. Present-day versions of these tests provide scores for four components, namely, the "traditional" language macro-skills of listening, reading, writing and speaking. The tests use direct, performance tasks for writing and speaking, and involve test materials that seek to simulate real-world tasks related to college and university.

The perceived need for testing has continued to grow. For example, in order to gain control over human migration flows, a jurisdiction may prioritise language proficiency as an important aspect of the employability of incoming migrants (Fulcher, 2010). However, rather than this decision leading to the development of tests for this new purpose, it has generally been the case that existing tests have been repurposed, despite the possibility that the test might be inappropriate or unvalidated for its new use.

Test providers tend to revise their public information about the kinds of organisations that "accept" the test, and the test enters new institutional and sociopolitical worlds. This has largely been the case in the Australian context, where, since the 1990s, the academic English language tests have done double duty at the gateway for tertiary study and skilled migration (see Frost & McNamara (2018) for a detailed overview of the use of tests for study and migration in Australia). The exception is the Occupational English Test (OET), which was purpose-built for health professionals (McNamara, 1990). While the use of academic English language tests for migration purposes may be practical, in that applicants who wish to migrate in order to study only have to do one test, it also allows the test purposes to merge in policy discourse and signals that they are viewed as interchangeable.

In Australia, major professional bodies, such as the Medical Board of Australia and Engineers Australia, are responsible for professional registration. In many cases, these bodies also act on behalf of the government as *assessing authorities* who assess the skills of applicants for *skilled migration* visas. Skilled migration is a migration category which applies only to defined *eligible skilled occupations* (e.g., telecommunications engineer), each of which is subject to intake quotas based on demand for each skill (Department of Home Affairs, 2021b). Up to 44,000 skilled independent visas have been issued annually in recent years (Gothe-Snape, 2018). The procedure for dealing with these large numbers of skilled migration applicants involves a system of awarding points, in which language test scores that are higher than the minimum requirement attract more points. This aspect foregrounds English language test scores in the competitive Australian migration system and indicates a high premium placed on English language skills. In addition to their role in the competitive application process, English language skills are embedded in legislation on general skilled migration along with health, character and other requirements to be met by applicants (Department of Home Affairs, 2021a).

Test mandates

Investigating how stakeholders view and represent the tests and test scores they use in their policies allows us insight into what they are expecting of a test, the meanings they ascribe to a score, and the impact they consider the test to have. We refer to this as an investigation of *mandate discourse* because it is concerned with the rationale for the use of a test, regardless of whether the rationale aligns with its designed purpose. Davidson and Lynch (2008) describe the test mandate as "the combination of forces which help to decide what will be tested and to shape the actual content of the test" (p. 77). Although this definition is primarily concerned with the motivations for the development of new test designs, a test mandate is an evolving concern which is often not the basis for a new test, but rather the basis for the use of an existing test to fulfil a newly arising societal or organisational need (see also Fulcher & Davidson, 2009).

Because discourse is dynamic, it can signal shifts in what users think a test is doing, without there being any shift in the characteristics of the test itself. This is important since it is in the nature of standards to be "ambitious for wider and wider acceptance" (Davies, 2008b, p. 439) or what we might call "mandate creep". The Common European Framework of Reference (CEFR; Council of Europe, 2001), for instance, has facilitated the proliferation of test instruments being accepted for certain purposes based on score equivalences alone, rather than on a consideration of the test construct, intended use and design (see Fulcher, 2004). As Davidson and Lynch (2008) point out, mandates "often remain unarticulated, implicit, and even murky" (p. 97). In the sociopolitical trajectory of academic English language tests described above, the articulation of the relationship between the original mandate (to determine readiness for English-medium universities) and the test's design and content may lose clarity as a test gains acceptance for a wider range of decisions. This murkiness does not simply result from single policy decisions; it occurs as a result of the needs of multiple parties interacting with the prevailing socioeconomic and political forces.

Mandate discourse occurs at two levels, as described by Gee (2014): first, as "little d" discourse, in communicating about tests through various modalities, which we present here via the perspectives of professional bodies and publicly available test information from the providers, and, second, as "big D" Discourse, in broader, historically formed, group-based patterns of understanding. We can witness this Discourse in the gradual repurposing of academic English tests for professional uses, and, even more

fundamentally, in the emergence of testing practices at the boundaries of a jurisdiction, or the boundaries of domains such as those of international English- medium education and skilled migration in the Australian context.

Test standards

The term *standards* has a number of meanings in the language testing literature (Davies, 2008a, 2008b). Here we use *standard/standards* to refer to threshold scores (also called *cut-scores*) required for professional registration and migration visas. This use of the term has a specific meaning, as a score that has been deemed to be a decision point. It also has a broader meaning in "big D" Discourse. The broader meaning of "tests" is encapsulated by McNamara and Shohamy (2008) who observe that tests are "associated with standards, objectivity and merit, and, in the context of immigration, are associated with productivity in the workplace and in society as a whole" (p. 89). This broader understanding of the objectivity, merit and productivity associated with tests is fed by a "standard language ideology", a belief in linguistic invariance, uniformity and correctness (Milroy, 2001). Adding a sociological layer: modern society is replete with standards of various kinds (traffic lights, bed sizes, shipping containers) which, for better or worse, enable "social and moral order" across distance and over time (Bowker & Star, 1999, p. 3; Busch, 2011). So, although we are focusing here on particular test standards in the form of test cut-scores, these instances cannot be dissociated from general understandings about what languages are and the modern drive to standardise things that inhabit the "big D" Discourse. These understandings make test standards possible.

Research study

In Australia, there are three professional bodies for accounting and one professional body each for engineering, nursing and medicine. In this study, we explore the meanings policy makers from these organisations ascribe (as score users) to test standards. We take an interpretative and inductive approach in which "the construction of meaning emphasizes the tacit knowledge surrounding a particular policy, its actors, and interpretive communities" (Moore & Wiley, 2015, p. 154). Our aim is not to evaluate organisations and their representatives, or test providers and their tests. On the contrary, the participants were invited to talk about their organisations' uses of language tests to us (we identified as language assessment researchers), and their participation arose out of a genuine concern for fair and informed uses of tests. Our aim is to understand how these score users view the role of tests and that of the specific test standards used in their professions so that we can better understand the uses of test scores in migration policy. Therefore, we contextualise these views in our subsequent discussion with other relevant information, particularly that available from test providers, who are key players in the interpretive community.

The interview data were drawn from two research projects which looked at the impact of English language tests for professional purposes (Knoch et al., 2016; Macqueen et al., 2013). In this study we focus on only the aspects of the interviews which show how these policy makers talk about the role of tests and test standards or cut-scores. As these interviews occurred in 2013-14, they do not necessarily reflect the current standards, procedures and views of the Australian federal government or the registration bodies; for this reason, we consider the sources in their historical context as evidence of uses of standards over time.

Participants

Eleven representatives from six Australian professional registration bodies took part in semi-structured interviews about their organisations' uses of English language tests in skilled migration and professional registration processes. The participants were from: accounting bodies (6 representatives), engineering body (3 representatives), medical body (1 representative), and nursing body (1 representative). Participants' roles in the organisations ranged from board chairpersons to representatives working on technical aspects of registration (e.g., checking certification and managing migration assessment). All participants had a clear interest in fair assessment, and in talking to language testing researchers, as indicated by their voluntary participation in research about their organisations' uses of English

language tests. Representatives are identified in the findings by the profession they represent, so, for example, the six accounting representatives interviewed are named Accounting R1 to R6.

Analysis

The interviews were carried out by the authors. They were approximately one hour each, and topics ranged from the specifics of test tasks to the uses of scores. The audio recordings were transcribed and then analysed in an iterative process (Miles et al., 2014). Initially, transcripts were annotated and sections of interviews which dealt specifically with the uses of language tests were identified. In this process, it became clear that the descriptions and rationales for test standards used by the various boards each had a historical element; the current test standards responded to a prior problem or an emerging challenge. In narrative analysis terms, these are what Barkhuizen (2020) calls *short stories*, which demonstrate how board members made sense of their use of test standards, often within the broader "story" of migration policy and shifts in assessment procedures and minimum scores. The narratives about test standards from each organisation provide points of comparison. These, we interpreted in themes, which we use to structure the findings. We do not mean to suggest that there is not much in common across the professional bodies, but we do mean to emphasize that the organisations face different challenges in relation to test use. Finally, in our discussion, we draw on other *symbolic artifacts* in the form of test provider information in dialogue with the views of the board members (Moore & Wiley, 2015).

Findings

The findings are organised topically based on representatives' perceptions of their organisations' uses of language tests, with similarities and contrasts drawn between the views of different participants.

Reducing risk and maintaining order

Language test standards were strongly associated with risk reduction and a concomitant maintenance of order through sorting on the basis of language proficiency in the workforce. For medicine, nursing and, to a lesser extent, engineering, risk reduction was the predominant concern. For accounting, the maintenance of consistency across the three professional accounting bodies, was a key concern.

For the medicine and nursing representatives, English language tests were seen as a first filter offering "a level of assurance about the basic competence in the English language" (Medicine R1). In these domains, English proficiency is explicitly linked to public safety: "registrants have to be able to speak English to a certain level to protect the public safety" (Nursing R1). Compounding this focus is the fact that Australian medical practice includes relatively isolated placements where the shortage of medical practitioners is acute in what are known as "Areas of Need" (Medical Board, 2020). This means that immigrant professionals may be practising in less-supported conditions than might be the case in other jurisdictions. Such placements highlight the interaction between risk and skill shortage. The decision, as the medical representative noted, may come down to the question of whether "for this community in this situation, this doctor is better than no doctor" (R1). For engineering, the safety issue arose in relation to communication about safety: Engineering R3 observed that part of the rationale for "a reasonably high standard of English" is that any ambiguity in understanding safety warnings and giving safety briefs are "very serious matters".

The assurance provided by a language standard was clear in relation to past skill shortages in the nursing profession, as the nursing representative describes:

> That risk is starting to decrease now as legislation's gone along. Certainly three, four, five years ago, that was a bit of a nightmare then, but the legislation is here, and the standards are here now. (Nursing R1)

Bringing order to some area of a social world is the job of standards (Busch, 2011) and, for nursing, we see that, along with risk reduction, there is also a sense that language test standards have helped to

bring order to the skilled migration assessment process.

For accounting bodies, the bringing of order was important in relation to the high volume of applicants, rather than a concern for risk:

> We don't have discretion. Based on volume, that's not something that we would be able to manage … in particular for ourselves, we have a higher market share than the other two [accounting] bodies – we wouldn't be able to manage it. None of the bodies would be able to. … when any assessment becomes discretionary, and it's all subjective, … it's hard to manage. (Accounting R2)

As the accounting representative emphasizes, an objective standard manages the complexity of dealing with large numbers of applicants. This concern reflects the fact that accounting had a much larger intake quota on the government skilled occupation list than the other professions, despite the fact that changes to the points system and IELTS requirements had led to a fall in the number of applicants (Accounting R4). Related to this is the fact that, for the accounting bodies, maintaining order was also tied more explicitly to competition in the global professional accreditation market.

Providing objectivity

Another valued characteristic of language tests across the professions was the perceived objectivity of language test measures – in particular, IELTS. We have already seen this in relation to the need for an objective measure to deal with the volume of applicants described above by Accounting R2. The perspective on objectivity from medicine was conjoined with the assurance sought in the use of English proficiency standards as a first filter: "It's important that we've got a basic level of English competency against some objective test" (Medicine R1).

The government's endorsement of language tests for migration assessment is important in the maintenance of objectivity, as Accounting R2 explains regarding the use of IELTS: "it's an expert test, something that's accepted". Similarly, the engineering body, in the process of determining how to assess the language proficiency of chartered engineer applicants, was introducing an IELTS requirement: "as a way of demonstrating its robustness and credibility, we're putting it up for third-party quality certification" (Engineering R2).

Providing consistency across governance structures

A further characteristic of test standards was consistency across jurisdictions or institutions. This was most apparent in the interviews with accounting bodies, each of which referred to the importance of a single standard for the three associations despite the fact that the three differed in terms of the types of accounting professionals and companies in their membership base, for example: "we all have to be aligned, all the three accounting bodies" (Accounting R2). Accounting professional bodies are also in competition globally, with each networked in different ways across national jurisdictions. For medicine, consistency was ensured across Australian states, when "in 2008 the National Medical Board agreed to a consistent English language testing policy around the country" (Medicine R1). For nursing, the issue of consistency in relation to medicine arose, following a report that showed pass rates for nurses at the first sitting of the English language test is not as high as for medical graduates (see Hawthorne, 2015).

Extrapolation of standards

Throughout the discourse, there was reference to anchor scores (mostly IELTS scores), which were used as reference points from which other score levels were extrapolated for other purposes. The individual professional bodies have each determined their score levels in relation to the government's minimum requirement for skilled migration, which was IELTS Band 6 at the time of the interviews. This standard has been referred to more recently in government literature as "competent English" and specified with equivalent scores in other commercial tests recognised as alternatives to IELTS (Department of Home Affairs, 2021a).

The accounting bodies had extrapolated a higher minimum from the government standard, as explained by Accounting R3: "the Department of Immigration have put general at 6, whereas we actually have it at 7 … and we all have to be aligned, all the accounting bodies … that's probably been our most recent challenge for candidates, because they're not quite reaching that 7". The difficulty of achieving IELTS Band 7 for the accounting applicants was observed by representatives from all three accounting bodies. Accounting R3 explained that she frequently processed applications with less than the minimum standard of IELTS Band 7 because international students can "either do IELTS Academic 7 or they can do their professional year". Entry to the professional year, known as Skilled Migration Internship Program Accounting (SMIPA), was set at IELTS Band 6 and one advantage of the program described by the professional bodies is that it "provides an opportunity to independently increase your English language proficiency" (CPA Australia et al., 2018). As Accounting R1 explained, "it's become a little more attractive for some people, as they've struggled to get IELTS 7".

The difficulty of the IELTS Band 7 minimum score was also perceived to lie in the accounting bodies' prescription of the Academic version of IELTS, a decision reportedly taken by all three accounting bodies "even though the Australian Government is happy to accept general tests [i.e., the IELTS General Training version]" (Accounting R3). Accounting representatives stressed the need for a high standard of academic ability, especially as the list of relevant occupations comprises senior positions such as financial controller, for which "Academic 7 is a must" (Accounting R4). A further extrapolation from the government language standard occurred beyond the professional bodies, as Accounting R1 explains: "there was some concern the past few years about Big Four accounting firms, in particular, not even considering international students for positions … now you'll see some of their websites that they will … Requirement? IELTS 8!" (see also Smith et al., 2016).

Thus, for accounting, the extrapolation of the government anchor standard of IELTS Band 6 (General Training version) is extended upward to the professional bodies' requirement of IELTS 7 and even to IELTS 8 for entry to higher-status employers, and from there, subsequently downward to IELTS 6 for the professional-year program, which typically leads to less senior roles for graduates (according to Accounting R1: "the majority are at low level, so book-keeping-type tasks"). The government IELTS 6 standard is also extrapolated from the General Training to the Academic version of the test, which comprises a reading sub-test in a more academic register and one different task of the two tasks in the writing sub-test.

The engineering body, by contrast, did not extrapolate from the government anchor standard for skilled migration, maintaining General Training IELTS 6 for this purpose. They had originally also seen no reason why experienced engineers, applying for chartered engineer status would "need a better level of communication than a graduate", since both may be required to carry out complex professional communication (Engineering R3). However, through their experience with lengthy written applications for chartered engineer status from people who had achieved IELTS Level 6, Engineers Australia had determined that IELTS 6 was an inadequate level (Engineering R3). As Engineering R1 explains, "we expect them to be a little bit better than your graduate engineer by the time they're seeking charter". Thus, they had extrapolated a score of IELTS 7 for chartered engineer status, an assessment process which certifies "a competent experienced engineer who could practise largely independently" (Engineering R3). Further, the IELTS score descriptions were to become influential in the body's professional communication standard, planned to "speak to level 7 as a guideline" (Engineering R2). Their process of understanding the meaning of IELTS 7 involved comparing samples of IELTS 7 writing with samples of narrative writing about professional experience by applicants who had obtained IELTS 7:

> Until we got the [published IELTS samples] DVD, we were I think naturally thinking: well, there's a piece of paper that says their writing's a 7, there's their writing, that must match; that must be what a 7 is. (Engineering R3)

Here, the engineering representatives provide a rare insight on the difficulty of determining and matching language ability across different language samples at the same gatekeeping moment – one

sample a lengthy professional narrative, the other a brief writing-test task.

Medicine and nursing representatives presented their bodies' language standard as an absolute point below which there was no assurance of adequate safety: "a person's got to be registered with an English language standard of 7 IELTS … that's the minimum standard … go away until you can prove you can do it and then once you can, then fine, we'll give you registration" (Nursing R1). Medicine R1 explained that applicants have to have "a medical degree … the required standard of English competence, and for almost everybody, you have to pass the AMC [Australian Medical Council] written exam, and if you haven't, [if] you don't approach the start- gate with those three … then you're rejected". Both representatives observed that the language standard was high, but necessarily so.

Separating language from professional competence

Language test standards were seen as distinct from the assessment of professional competence. This undoubtedly arises from the fact that Australian law requires the separate assessment of language skills (McNamara, 1996). However, all four professions include communication within their professional attributes. Thus, language test standards may be viewed either as tapping into a professional competency to some extent, i.e., *professional communication*, or as a completely separable skill, *language proficiency*. The latter view is expressed by Medicine R1:

> We don't think it's testing clinical communication skills … that's a whole lot of stuff … there are many native English speakers who don't have good communication skills in a clinical context.

The medicine representative explained that language tests act as a kind of first filter of the applicant pool, a necessary requirement along with other layers of professional assessment, for example, the written Australian Medical Council exam. The separation of medical communication skills and language skills is also clear in the expectations of what a language test construct should reasonably encompass:

> If testing is congruent with practice that's terrific, but we shouldn't be relying on that as the method for saying [that] these people will be good clinical communicators and culturally competent … because that's just not reasonable. (Medicine R1)

In this view, the language test is not an indication of professional communication skills although some overlap between the test and practice may be a positive aspect. Similarly, the nursing representative (R1) suggested that employers may have "more confidence" in an "occupationally appropriate" test.

The interviews also revealed the dynamism around the test instrument (at that time, mainly IELTS) for the different bodies in the relationship to professional communication. Medicine R1, for example, pointed to the ongoing need as a regulatory authority to "reflect on what you're doing" to make sure "it's contemporary" in relation to "a bigger issue around the way in which communication is changing". Accounting was experiencing a change in professional boundaries because "it's not about numbers anymore … it's not so narrow anymore" (Accounting R2). As we have seen, the main concern for engineering was related to how to relate professional communication as represented in professional assessments for chartered engineer status to language test samples. This concern arose primarily from the changing population of applicants for chartered engineer status.

<div align="center">

Discussion

</div>

Test standards and trust

We propose that the concepts arising in the discourse presented above are underpinned fundamentally by a sociological conceptualisation of trust: trust in the test standard (an abstract conceptualisation, usually called "IELTS 6" in the talk of the Australian professional body representatives, for example), as well as trust in the standard language (also an abstract idealisation called "the English language"). Themes in the findings such as *order, risk reduction, consistency, objectivity* and *extrapolation* all invoke trust in the test standard. Trust is developed through

understanding something as being predictable in the face of risk (Lewis & Weigert, 1985).

Language test standards offer a trusted sorting apparatus for highly complex phenomena (language and language ability) in increasingly complex social and institutional circumstances. Luhmann (1979/2017) has argued that trust is a mechanism for reducing complexity (p. 33). Modern society is infused with what he refers to as "system trust" which enables dealings between people who are otherwise unconnected (p. 61), for example, a Sri Lankan nurse and an Australian aged-care provider. The nexus of modern migration trends, diverse languages, and diverse educational sources of professional competence is certainly a complex amalgam into which familiar and trusted objects such as language test standards offer welcome solutions to the impossible task of predicting the dynamic behaviour of individuals moving from a specific work context in one professional jurisdiction to another. In such complex circumstances, a sociopolitical need for trusted test standards to manage the imagined smooth operation of the workforce readily arises. In this way, language tests embody what Luhmann (1979/2017, p. 17) describes as the future- oriented nature of trust: from test scores, at best, we can infer someone's language proficiency in a future workplace; scores are incapable of assuring it. As Lewis and Weigert (1985) phrase it, "trust begins where prediction ends" (p. 976). The social operation of extrapolating someone's performance in an actual workplace from a test score therefore requires an "ambitious leap" (Kane, 2013, p. 28).

Trusted test standards not only reduce complexity, they are tied to the central concern of risk in professional practice (Knoch & Macqueen, 2020). In the discourse presented here, different kinds of risks were raised. Medicine, nursing and engineering board representatives express concern about the risks of harm resulting from inadequate language skills. Risk (and safety) have been the focus of public consultations on the professional and language requirements for medical and other healthcare professionals (Pill & Harding, 2013). Accounting board representatives, on the other hand, do not suggest any relationship between language standards of accounting professionals and risk to the public. Their concern (and resulting placement of trust in the test standards) is more with the risk of having inconsistent standards across the separate accounting bodies and management of a large volume of applicants.

Familiarity and extrapolation

A precondition for trust is familiarity (Luhmann, 1979/2017, p. 19). Originally developed for academic purposes, IELTS has since expanded its advertised uses to migration and work, specifically for professional registration. Bodies are advised to determine a suitable band score and test version (Academic or General Training) (IELTS, 2021). IELTS has been the preferred choice of Australian universities and the Australian Government since the 1990s (O'Loughlin, 2011). The IELTS score required for immigration has gradually risen over the years, starting from a score of Band 4.5 in 2009 for a temporary work visa, then raised to 5 accompanying a reduced need for labour (Fulcher, 2010). The test score commonly referred to as "IELTS 6", and its associated description of "competent English", has since been reified in Australian government discourse, where it signifies the minimum level of English for participation in "skilled" occupations (Department of Home Affairs, 2021a). More recently, IELTS Band 6 has been proposed as a standard for Australian citizenship, with robust debate in parliament as to the meaning of "IELTS 6" on the General Training version of the IELTS test (Macqueen & Ryan, 2019). As test standards are extrapolated to different uses, we can also witness "mandate creep", where test standards offer seductively simple solutions to complex perceived problems.

In the professional bodies' discourse, we saw IELTS Band 6 had become an abstract anchor standard which was extrapolated to various uses. These extrapolations, and some subsequent ones, are shown in Figure 1.
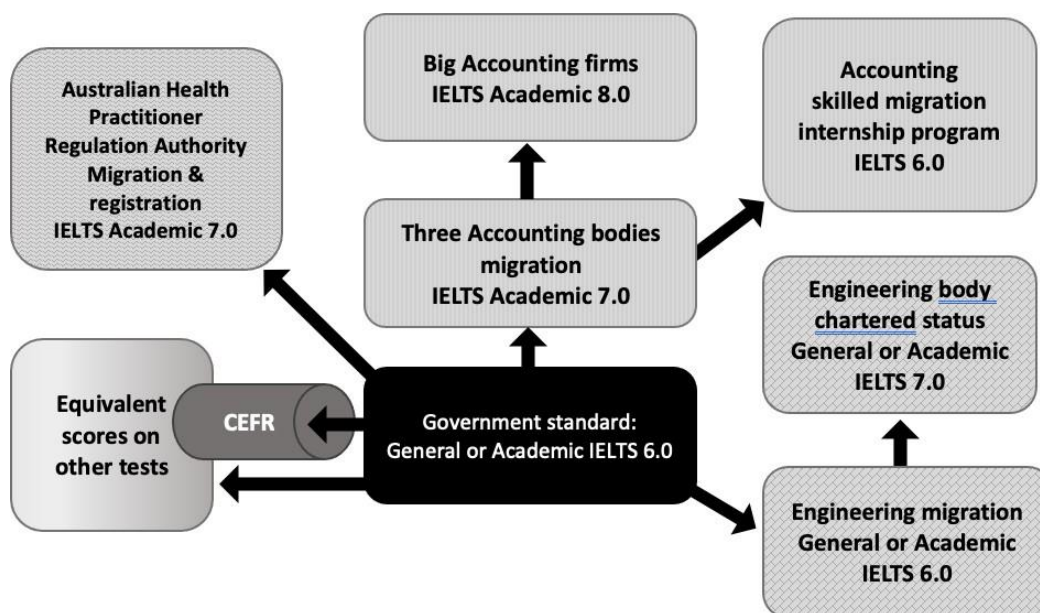
Figure 1: Extrapolations from the government anchor standard of IELTS 6

Moving forward from the historical context of the research interviews (in 2013-14), we observe that the standard has subsequently extrapolated further. All professional associations have since published equivalence tables which now include other language test standards. The last 10 years have seen a great deal of aligning activity between commercial tests. One way for this to happen is for tests to be aligned to each other, with the advantage that a less familiar test standard can gain a foothold in a market by being linked to an already trusted one (e.g., Educational Testing Service, 2010). Another way is for test providers to align their instruments to the CEFR, using this as a parent standard. The CEFR has facilitated equivalence tables which set out equivalences via simple, non-descriptive numerical or alphabetical forms (scores, bands, levels, etc.). These score equivalences engender the understanding that the tests are equivalent without consideration of their constructs (i.e., what is being tested in terms of skills, language and tasks). Arguably, scores are great conductors of trust due to their objective appearance (Porter, 1996). It seems likely that the trust witnessed in these findings in the use of IELTS 6 as an anchor standard is the propellant which enables extrapolation via equivalent scores to other tests. Through its use in Australian policy contexts over 20 years or more, IELTS 6 has "sunk into" the infrastructure (Star & Ruhleder, 1996, p. 113). The experience of the engineering representatives who were trying to understand the test standard from first principles, by looking at test samples and samples of their own professional assessments, is illuminating in that it shows the challenge of deriving the meaning of test scores in a qualitative, relational sense, even for these highly experienced assessors of professional competence.

Once gained, trust is extended to scores above and below the required level on the assumption that the language abilities vary just as numbers do – they are higher or lower in the same way that water from one sample will be much the same if there is more or less of it in a glass. There is no consideration that the scores might indicate a qualitatively different language sample (and underlying ability), despite the fact that the parent standard, the CEFR, distinguishes language ability levels entirely on a qualitative basis. It is tempting for policy makers to assume that score equivalence is the same as fit-for-purpose equivalence, despite the likelihood that the array of tests in an equivalence table represent significantly different operationalisations of the same stated construct, for example, English language proficiency (Knoch & Macqueen, 2020).

Reproduction of trust

Thus, the familiarity required to trust a test arises, at least in part, from its presence and use in powerful places, rather than from a more detailed knowledge of the instrument or what it measures. This reflects the fact that acts of trust encourage others to place trust in similar ways (Lewis & Weigert, 1985), and it explains how a principle of "use begets use" appears to operate for test standards. The

influence of the trust- placement of others in a test object is strong enough to override distrust, created, for example, by deceptive use of test reports, which took up a great deal of time for one professional representative. Other negative anecdotes – such as groups of students travelling around the country to repeat IELTS tests to achieve the necessary score (Accounting R1), and the concern that nursing cohorts score lower than doctor cohorts (Nursing R1) – did not appear to diminish trust in the test or its role in migration sorting overall. Possibly the *system trust* achieved by the test over time enables these reasons for distrust to be eclipsed by the sheer complexity of the policy circumstances being managed by the test and the difficulties foreseen in replacing it with another mechanism (Luhmann, 1979/2017, p. 62).

The centrality of trust is a familiar feature of marketing discourse generally, and for language tests this is also the case. All English language tests currently accepted by the Australian Government use "trust" as a marketing mechanism. For examples, we turn to the websites promoting these tests (emphasis added):

> *James Shipton, Head IELTS at the British Council, said: "The continued growth of IELTS that we've seen around the world is testament to the popularity of the test with test takers, and to the **trust** that organisations place in IELTS to provide scores that are a reliable indicator of a person's ability to communicate in English."* https://www.ielts.org/news/2017/ielts-numbers-rise-to-three-million-a-year

> *Team up with IELTS to give your clients access to exclusive preparation tools, free study tools, and promote your business' status of **trust** and reliability by using the official IELTS logo.* https://ielts.com.au/ielts-referral-program/

> *More than 11,500 institutions in over 160 countries **trust** the TOEFL test to help them make informed, confident decisions about the English-language skills of their applicants.* https://www.ets.org/toefl/score-users

> *Accepted by the Australian, New Zealand and UK governments | **Trusted** by over 3,000 universities and colleges.* https://pearsonpte.com

> *OET is recognised and **trusted** by healthcare boards and councils in the UK, the US, Canada, Ireland, Australia, New Zealand, Ukraine, Dubai, Singapore and more.* https://www.occupationalenglishtest.org/organisations/government-healthcare- boards-councils/

> ***Trusted** results for important decisions |More than 25,000 organisations in 130 countries around the world rely on our secure exams and tests as proof of English language ability* https://www.cambridgeenglish.org/why-choose-us/global- recognition/organisations/

Language tests play a critical role in converting language-as-a-commodity to a "currency" (i.e., recognised test scores) for use in a neoliberal economic system. Of course, decades of research in second language acquisition and sociolinguistics on variables such as first and second language exposure/access attest to the complexity and potential injustice of such a conversion. Ricento (2012) has argued that language policy discussions on the role of English as a global language have not fully grasped the influence of neoliberal economic values and priorities in the policies of states and international organisations (p. 32). He highlights the fact that in economic terms, English proficiency is tied to particular advanced educational credentials as the basis of a knowledge economy. In this vein, Lo Bianco (2021) and others have developed the notion of "linguistic entrepreneurship" as a consequence of human capital – the imperative to develop language skills to compete in the market for employment and for other life chances. What we have tried to demonstrate here is that trust is a key ingredient in the market forces that sustain the use and spread of language test standards as part of the boundary infrastructure of a knowledge economy. While trust may be well-placed, it must not be taken as evidence of test suitability, either by testing agencies or test users.

The language testing industry (providers and developers), language testing researchers, and score users (governments, institutions, organisations, etc.) have a responsibility to consider the market forces that (1) entrench test standards and (2) nurture their proliferation and extrapolation. The effect of these forces is that they enable understandings, and then uses, of tests and particular score thresholds ("test standards") to be generated by marketing efforts and reputation, and not on the basis of construct-relevance or evidence of suitability. Kane (2013) has argued strongly that in approaching test validity, equal consideration should be given to *interpretations* of test scores and *uses* of test scores because "arguments for the appropriateness of a score use typically lean heavily on the relevance of score interpretations" (p. 2). What we aimed to show here is that interpretations of test standards, i.e., particular scores that are put to use in policy, occur in both "big D" Discourse and "small d" discourse (Gee, 2014). That a test can be a product for purchase, or that a test might be used to sort people in and out of geographical spaces are naturalised understandings in modern societies. That is, they manifest in "Big D" Discourse. These kinds of widely held understandings lay the foundations for related, common understandings such as the truth-value of a particular test standard in the context of the neoliberal economic orders of many "standard language cultures" (Milroy, 1999, p. 18). Once a test standard has become absorbed in a policy infrastructure, mandates attach to it like vine tendrils seeking a trellis.

Practical implications

Interrogating these understandings has some practical implications. First, language test providers must evaluate (their) communications about particular test instruments critically. This is the level of "small d" discourse, and it includes the nature of the specific information provided to clients/stakeholders and their interpretations of the test standard. Interpretations as to the suitability of a test may be generated via an entrenched test score, a parent standard such as the CEFR, an alignment to another test, a use by a similar organisation, use in a parallel policy in another jurisdiction, or some other means. Consideration should also be given to the interplay between the stated construct and other kinds of information advertised. One facilitative mechanism for mandate creep is the common practice of listing corporations and organisations who accept a test, irrespective of whether or not listed institution/organisation use is aligned with the designed use of the test (see also Knoch & Macqueen, 2020). Rather than quietly allowing uses to stretch beyond the appropriate remit of a test design, testing agencies could offer clarity for score users by being explicit about uses arising in the client base which are not within the intended scope of the instrument.

Second, policy makers are advised that their choice of language test has an impact on the language abilities of most applicants through the kinds of test preparation activities the test generates. The fact that language tests tend to come under the banners of "compliance" or "accreditation" means that their potential for developmental effects on a workforce (e.g., through learning domain-relevant language) are less obvious. A more holistic skilled migration policy might pay attention to the development of relevant language abilities prior to entering the workplace, and not just to their measurement.

Finally, language testing specialists can be more "policy responsive" (see Elder, this issue), by improving their understanding of how and why tests are used, both within and beyond the designed uses. Mandate discourse can be examined through scores users' perceptions of tests and test standards, as we have done here. Other perspectives on the social lives of tests that warrant attention are test-takers' experiences and perceptions of tests (see Frost, this issue), and test providers' discourse about their tests. These kinds of explorations can lay foundations for more effective communications between language testing specialists and score users about the importance of ensuring that test uses are congruent with test constructs. A greater understanding of the kinds of complexities involved in determining policy settings will go some way to facilitating "policy responsible" conversations and interventions (see Elder, this issue).

Conclusion

English language proficiency is a deciding factor in the life opportunities of many thousands of

applicants for Australian skilled migration every year. This paper has focused on the perspectives of professional bodies who use English language tests in skilled migration decisions. The policy narratives around the use of test scores describe the need to manage large numbers of applicants, to assure a level of English proficiency for high-risk professional communication, to provide an objective assessment that is separate from any assessment of professional competence and to maintain consistency of standards across bodies.

The sociological concept of *trust* (Lewis & Weigert, 1985; Luhmann, 1979/2017) offers some explanatory power. Trust is fundamentally an exercise in the reduction of complexity, including unpredictabilities such as public risk and diverse applicant populations. By meeting the need for a solution to the complex reality of migration selection processes, professional communication and large numbers of applicants (among other factors), tests and their standards become trusted policy objects, using the apparently simple semiotics of numerical scores to communicate across regulatory boundaries. Standardised language tests, for all their mechanistic complexity, are tools of rapid and even brutal simplification. They render an "empirically intricate reality deceptively straightforward" for the purpose of comparison (Stevens, 2008, p. 102).

Standards, such as test standards, tend to become entrenched, invisible, and nested in other standards, creating inertia and preventing change (Bowker & Star, 1999). Because adaptation happens around them over time, it is in the taken-for-grantedness of abstract standards such as "IELTS 6" that their power resides (Busch, 2011). Fulcher and Davidson (2007) refer to these entrenched standards as "iconic scores" and warn that score meaning does change when tests themselves change, for example, with a change in technology (pp. 92-93). In this study, we see the entrenchment of scores from another angle, where trusted test scores might be extended to new uses, or other tests may come into use by virtue of correlational studies with trusted scores but not necessarily through their construct relevance. As Luhmann (1979/2017) observes, "Trust is only possible in a familiar world; it needs history as a reliable background" (p. 23). However, the fact that a test standard is trusted does not mean it could not be improved, since "there is no natural law that the best standard shall win" (Bowker & Star, 1999, p. 14). As a test is absorbed into the system, its influence extends beyond more immediately apparent impacts to system-wide interactions, such as score extrapolations. We contest that a test should be worthy of the trust placed in it, as well as potentially beneficial in the language skills it develops in its test-taker population.

References

Barkhuizen, G. (2020). Core dimensions of narrative inquiry. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 188- 198). Routledge. https://doi.org/10.4324/9780367824471-17

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. The MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Busch, L. (2011). *Standards: Recipes for reality*. The MIT Press. https://doi.org/10.7551/mitpress/8962.001.0001

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* https://rm.coe.int/168045b15e

CPA Australia, Chartered Accountants, & Institute of Public Accountants. (2018). *Accounting professional year program: Participant frequently asked questions*. https://www.cpaaustralia.com.au/-/media/corporate/allfiles/document/become/migration-assessment/accounting-pyp-faqs.pdf?la=en&rev=9fc70d21188147f084020673f41cedcd

Davidson, F., & Lynch, B. K. (2008). *Testcraft: A teacher's guide to writing and using language test specifications*. Yale University Press.

Davies, A. (2008a). Accountability and standards. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 483-494). John Wiley & Sons. https://doi.org/10.1002/9780470694138.ch34

Davies, A. (2008b). Ethics, professionalism, rights and codes. In N. H. Hornberger & E. Shohamy (Eds.), *Encyclopedia of language and education, Volume 7: Language testing and assessment* (pp. 2555-2569). Springer. https://doi.org/10.1007/978-0- 387-30424-3_191

Department of Home Affairs. (2021a, September 18). *Meeting our requirements: English language*. https://immi.homeaffairs.gov.au/help-support/meeting-our- requirements/english-language

Department of Home Affairs. (2021b, September 19). *Skilled occupation list*. https://immi.homeaffairs.gov.au/visas/working-in-australia/skill-occupation- list

Educational Testing Service. (2010). *Linking TOEFL iBT scores to IELTS scores: A research report*. Educational Testing Service. https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Frost, K., & McNamara, T. (2018). Language tests, language policy, and citizenship. In J. W. Tollefson & M. Pérez-Milans (Eds.), *The Oxford handbook of language policy and planning*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190458898.001.0001

Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly, 1*(4), 253-266. https://doi.org/10.1207/s15434311laq0104_4

Fulcher, G. (2010). *Practical language testing*. Hodder Education.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge. https://doi.org/10.4324/9780203449066

Fulcher, G., & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing, 26*(1), 123-144. https://doi.org/10.1177/0265532208097339

Gee, J. P. (2014). *How to do discourse analysis: A toolkit* (2nd ed.). Routledge. https://doi.org/10.4324/9781315819662

Gothe-Snape, J. (2018). *Government's immigration tweak sees overseas Asians out, integrated Kiwis in*. https://www.abc.net.au/news/2018-04-13/how-dutton-slowed-immigration/96466021

Hawthorne, L. (2015). *International health workforce mobility and its implications for the Western Pacific region. Final report. Commissioned by the World Health Organization*. Centre for Health Policy, University of Melbourne. https://core.ac.uk/download/pdf/155015798.pdf

IELTS. (2021, April 13). *IELTS for work*. British Council, IDP: IELTS Australia, and Cambridge Assessment English. https://www.ielts.org/about-ielts/ielts-for- work

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Knoch, U., & Macqueen, S. (2020). *Assessing English for professional purposes*. Routledge. https://doi.org/10.4324/9780429340383

Knoch, U., May, L., Macqueen, S., Pill, J., & Storch, N. (2016). *Transitioning from university to the workplace: Stakeholder perceptions of academic and professional writing demands* (IELTS Research Reports Online Series). https://www.ielts.org/-/media/research-

reports/ielts_online_rr_2016-1.ashx

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63*(4), 967-985. https://doi.org/doi.org/10.2307/2578601

Lo Bianco, J. (2021). The discourse of the edge: Marginal advantage, positioning and linguistic entrepreneurship. *Multilingua, 40*(2), 261–275. https://doi.org/10.1515/multi-2020-0188

Luhmann, N. (1979/2017). *Trust and power* (H. Davis, J. Raffan, K. Rooney, M. King, & C. Morgner, Trans.; English ed.). Polity.

Macqueen, S., Pill, J., Elder, C., & Knoch, U. (2013). *Investigating the test impact of the OET: A qualitative study of stakeholder perceptions of test relevance and efficacy*. University of Melbourne, Language Testing Research Centre.

Macqueen, S., & Ryan, K. (2019). Test mandate discourse: Debating the role of language tests in citizenship. In C. Roever & G. Wigglesworth (Eds.), *Social perspectives on language testing: Papers in honour of Tim McNamara* (pp. 55-71). Peter Lang.

McNamara, T. (1990). *Assessing the second language proficiency of health professionals* [Doctoral dissertation, University of Melbourne]. https://minerva-access.unimelb.edu.au/handle/11343/35332

McNamara, T. (1996). *Measuring second language performance*. Longman.

McNamara, T. (2012). Language assessments as shibboleths: A poststructuralist perspective. *Applied Linguistics, 33*(5), 564-581. https://doi.org/10.1093/applin/ams052

McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics, 18*(1), 89-95. https://doi.org/10.1111/j.1473- 4192.2008.00191.x

Medical Board - Australian Health Practitioner Regulation Agency. (2020, September18). *Area of need*. https://www.medicalboard.gov.au/registration/types/limited- registration/area-of-need.aspx

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Milroy, J. (1999). The consequences of standardisation in descriptive linguistics. In T. Bex & R. J. Watts (Eds.), *Standard English: the widening debate* (pp. 16-39). Routledge.

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics, 5*(4), 530-555. https://doi.org/10.1111/1467-9481.00163

Moore, S. C. K., & Wiley, T. G. (2015). Interpretive policy analysis for language policy. In F. M. Hult & D. C. Johnson (Eds.), *Research methods in language policy and planning: A practical guide* (pp. 152-165). John Wiley & Sons. https://doi.org/10.1002/9781118340349.ch14

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*(2), 146-160. https://doi.org/10.1080/15434303.2011.564698

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381-402. https://doi.org/10.1177/0265532213480337

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press. https://doi.org/10.1515/9781400821617

Ricento, T. (2012). Political economy and English as a "global" language. *Critical Multilingualism Studies, 1*(1), 31-56.

Smith, B., Maguire, W., & Han, H. H. (2016). Generic skills in accounting: Perspectives of Chinese postgraduate students. *Accounting & Finance, 5*(2), 535-559. https://doi.org/10.1111/acfi.12219

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111-134. https://doi.org/10.1287/isre.7.1.111

Stevens, M. L. (2008). Culture and education. *The ANNALS of the American Academy of Political and Social Science, 619*(1), 97-113. https://doi.org/0.1177/0002716208320043