

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

Beyond dichotomies: Unravelling the complexities of international standardised English language testing

Luke Harding

Abstract

This article discusses the issues raised in O’Sullivan’s and Hashimura et al.’s Point-Counterpoint articles with the aim of introducing new angles on the topic of international standardised English language testing and opening the debate further. In the first section, I provide summaries of the positions taken by O’Sullivan and Hashimura et al., contrasting the “tests as a social good” and “critical language testing” paradigms. Second, I explore international standardised English language testing as a site of complexity, focusing on the nature of this complexity in conceptualising language testing as a field, language tests (as social objects), and language testing organisations. Third, I address shifts towards measuring innovative constructs and the prospect of greater localisation. Finally, I conclude with a call for research that explores complexity in greater depth with a view to change.

Key words: Language testing; standardised testing; construct; washback; English as a Lingua Franca

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

Introduction

A good debate provides an opportunity to explore two sides of a provocative issue.

However, the traditional debate format also encourages participants to present their views emphatically, to persuade the audience of a particular viewpoint set out with certainty. In committing those positions to writing, the result is often a set of stark

dichotomies. We see those dichotomies in O’Sullivan’s and Hashimura et al.’s point-counterpoint articles; two contrasting perspectives on the same question: are

international standardised English language tests beneficial to learners and teachers?

The points raised are thought-provoking and, often, refreshing. I agree with O’Sullivan that the lack of research into the potential benefits of international standardised

language tests is “shocking” given their influence and proliferation. I also agree with

Hashimura et al. that, “We are, all of us, inside it”; for many working in English language education, international standardised testing touches every corner of teaching and

learning, from the look and feel of materials, syllabus and curriculum goals, classroom assessments, and the shorthand used to describe language ability (6.5; B2; 94). A

debate of this kind is therefore vital and illuminating.

In this article, first I summarise and characterise the key arguments put forward by O’Sullivan and Hashimura et al.; second, I consider language testing as a site of complexity, where the interrogation of monolithic framings could lead to more revealing insights about language testing as a field, language tests, and language testing organisations. Third, I address the potential solutions offered by innovative constructs

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

and localisation. Finally, I will conclude with a reflection and brief research agenda. My aim throughout is to question assumptions, introduce different perspectives, and open the debate.

Characterising the arguments

The responses naturally reflect different vantage points: O’Sullivan – a language testing professional – presents a view of standardised language testing as a social good.

O’Sullivan provides a classic defence of standardised testing based on meritocratic principles, a key theme in many justifications of large-scale, standardised testing practices (see Fulcher, 2015). Within this view, standardised testing is seen as the most objective and fair form of assessment, minimising bias through consistency and rigorous quality control. As such, these tests are viewed as drivers of social mobility, providing equal opportunity for all regardless of wealth, connections or other advantages unrelated to talent or ability. In the landscape of international education, the tests function as passports – “opening doors to global opportunities” (O’Sullivan, 2026, p. [insert page number]) – by providing evidence of language proficiency in a format (the recognisable test score) that functions as a trusted common currency.

Hashimura et al. write from the perspective of a group of educators and researchers who have had first-hand experience as test-takers or working within the testing industry. While they acknowledge that international standardised language tests are “here to stay”, they argue “not to abolish assessment but to release it from the reductive commitments of overstandardisation” (Hashimura et al., 2026, p. [insert page

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

number]). Their response is firmly rooted in the tradition of critical language testing, an area of study where tests are understood as instruments of power, with potentially harmful consequences for learners, teachers, and wider society (see Lynch, 2001; McNamara, 2011; Shohamy, 2001). Hashimura et al. point to the uneven global economic conditions that impede access to the kinds of opportunities international standardised language tests offer, and characterise language test providers' provision of free resources as *edu-washing*: a means to appear benevolent while perpetuating the reach and influence of their own products. Hashimura et al. also argue that standardisation can lead to harmful lack of representation in test content, resulting in raciolinguistic bias permeating test constructs.

Interrogating monoliths

The dichotomies that emerge when these articles are read side-by-side – meritocracy versus lack of access, standardisation versus authenticity, mobility versus gatekeeping – point to key tensions that remain unresolved. These tensions are, however, not new to the field, and in fact are reminiscent of an ongoing negotiation within language testing and assessment over several decades. In this sense, it is important to note that the academic field of language testing is not monolithic. The two broad paradigms that underpin the two articles – which I will loosely describe as the “tests as a social good” paradigm and the “critical language testing” paradigm – co-exist, finding expression at conferences, in the pages of journals, and in the syllabi of language testing and assessment courses. While the “tests as a social good” approach may function as a

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

dominant paradigm, underpinning practices and reassuring stakeholders, critical perspectives continue to have an influence on mainstream thinking and, in some cases, changes to practice. Progress that has been made in language testing and assessment, shifting away from some of the more extreme forms of standardisation (through more communicatively-oriented tasks) and the development of an extensive research focus on washback and impact (see Allen, 2025), might be attributable to these two paradigms remaining in close dialogue. This duality suggests that language testing, as a field, is a site of complexity: the intertwining of competing paradigms and the dynamic relationship between these apparently dichotomous worldviews is a feature of how the field is organised.

There are aspects of both articles that might also benefit from a similar kind of complexification. In the case of O’Sullivan, international standardised language tests are often referred to as a monolithic whole, for example, “International standardized tests, with their focus on communicative competence and real-world skills, can drive positive washback ...” (p. [insert page number]). While many tests of this kind are premised on a theory of communicative language ability, the degree to which this filters down into a set of rich, communicative tasks varies. Aligning a test to the CEFR does not guarantee that the communicative approach that underpins the CEFR is also driving considerations of test design. In this sense, the only meaningful unit of analysis in considering the “communicativeness” and potential impact of international standardised language tests is to consider the test itself, separately, whether that be IELTS, TOEFL iBT, Duolingo English Test, Pearson Test of English Academic, Cambridge B2 First, LanguageCert, or one of the multiple other language proficiency tests that are

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

used across international markets. Each test has its own unique construct, characteristics, research base, and format. Stakeholders should be encouraged not to think at the aggregate level, but to scrutinise the particular strengths and weaknesses of each test, in order to evaluate its quality and fitness for purpose. Talking of international standardised language tests as a monolithic whole is likely to obscure rather than reveal important insights about the nature of language testing.

Hashimura et al. also approach language testing as a monolithic entity but from a different direction, making metonymic reference to “Princeton and London” and describing the “testing-industrial complex” (p. [insert page number]). On the former, it is worth noting that centres of gravity are shifting in the international language testing landscape. TOEFL and IELTS may no longer have the same kind of duopoly they once had, and the Duolingo English Test in particular is exerting a disruptive influence on the field through its capacity to deliver tests at home, with short timings, and at a low cost. Rather than functioning as bastions of concentrated wealth, over the past two years the British Council and Educational Testing Service have experienced significant financial strain leading to job losses (Knox, 2024; Nelson, 2025). This is not to say that the critiques levelled at international tests – “funnelling gains upward from candidates in lower-resourced contexts” (p. [insert page number]) – do not remain valid. But the critique will need to be adapted to take into account newer “players” and the novel impacts of AI-driven language testing on economic conditions and job market health across the wider language testing ecosystem.

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

In addition, the testing-industrial complex may well exist as a general force shaping English language education, but there is additional complexity when we look more closely at the personnel within language testing organisations. On the one hand, these organisations employ researchers, test developers, and international networks of item writers and raters, who may have deep and critical understandings of both the benefits and limitations of large scale testing; however, those same organisations include business and marketing divisions, which may also play a significant role in decisions around what is and is not operationally feasible. In my experience observing at close hand a range of testing organisations, the internal struggle between academic-research and business-oriented divisions is often the most important factor in shaping various aspects of test design, delivery, and innovation. This seems important to note: we should be intensely curious about the inner workings of these organisations. Understanding the internal structures of language testing providers has often been overlooked (except for Spolsky's [1995] *Measured Words*, which is essential reading for anyone interested in the topic). A richer sociology of the organisations that produce these tests would reveal fascinating insights about why things are the way they are (and not as perfect as we want them to be).

Constructs and localisation

In any aspect of language testing, it is always important to return to the construct: the fundamental ability that we want to measure. Both articles signal, in their own way, the vital role of getting the construct right. In the case of Hashimura et al., standardisation

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

is seen as “flattening” the construct – an evocative image – leading to a native-centric approach, whether overt or covert, in the constructs of international standardised language tests. O’Sullivan, too, nods towards the importance of construct when he notes, in the context of a discussion about positive washback, “of course, all of this is contingent on test providers developing ISTs [international standardised tests] that are demonstrably focused on contextually meaningful language constructs” (p. [insert page number]).

This leads to the inevitable questions: what are contextually meaningful language constructs? And how can they best be assessed in ways that promote and recognise the rich and diverse sociolinguistic reality of many language use domains? On this point, I have strong sympathies with the position adopted by Hashimura et al.; moving beyond native-centric constructs and recognising the global realities of English as a Lingua Franca in language testing have been driving forces for my own research (e.g. Harding & McNamara, 2018). And they are right to identify that a key constraint in implementing these newer conceptualisations is standardisation. If we shift to a view of language proficiency as dynamic, highly-contextualised, and fluid – ideas that have been advanced by Canagarajah (2006), Leung (2005), and others – problems emerge in trying to measure consistently and fairly across different candidates, with different interlocutors, in different contexts. These are problems, however, that language testers are currently trying to solve through research on assessing lingua franca competence, integrating Global Englishes perspectives, and reframing issues of representation through the lens of EDI (Equality, Diversity and Inclusion). While slow-going, there is encouraging progress underway.

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

I would propose, however, that the problem is not only with standardisation. A wider issue is the use of the same test across multiple contexts, where a test may experience “mission creep”, gradually being used for a range of purposes for which it was not necessarily designed. In such cases, the “international” nature of international standardised tests becomes problematic. As a case in point, a study by Iliovits et al., (2022) demonstrated that at a university in Lebanon, where TOEFL iBT was often used for admissions, a close analysis of the target language use domain (observations and recordings of classroom interactions) revealed that translanguaging and use of English as a lingua franca strategies were salient features. These features are not captured in the speaking constructs of tests such as TOEFL iBT, partly because the original domain analysis to support test design was based on the norms of US universities.

One solution for this problem is a shift towards greater localisation of language testing. The creation of tests that are more sensitive to context and have greater potential to capture what matters to stakeholders in their environments. To achieve this, however, requires resources, training and ownership at the local level. And here we come back to the *Realpolitik* of international standardised language testing: outsourcing decisions around language proficiency to large organisations with in-house knowledge and existing products will often be more appealing than developing a new instrument with all the time, money and effort required. As Iliovits et al. (2022) noted, creating contextualised, local tests that involve richer constructs is a surmountable conceptual challenge – it can be done and the frameworks exist to do it. The prime challenge relates to resource, time, and expertise.

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

Conclusion

Ultimately, O’Sullivan paints a picture of the impact of international standardised language tests that makes important points about meritocratic principles, but may more accurately describe a philosophical ideal rather than the reality experienced by the learners and teachers who engage with such tests. Meanwhile, there are no easy solutions to the problems identified by Hashimura et al., particularly as international standardised language tests occupy a multifaceted role in the landscape of English language education. Focusing on the influence of language testing organisations is warranted, but there also needs to be space for the consideration of wider social dimensions of test use, including the roles of prestige, of ritual, or of competition that may also drive demand for these tests in particular contexts. We require, therefore, a research agenda that moves beyond dichotomies and begins to explore the complexities and contradictions of language testing as a field, of language tests as social objects, of language test providers as multifaceted organisations, and of language test users as individuals with decision-making agency. Unravelling this complexity will lead to richer insights about why things are the way they are, and how we might go about making changes for the better.

References

Allen, D. (Ed.). (2025). *Washback research in language assessment: Fundamentals and contexts*. Routledge.

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*.
Routledge.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly: An International Journal*, 3(3), 229-242.

https://doi.org/10.1207/s15434311laq0303_1

Harding, L., & McNamara, T. F. (2018). Language assessment: The challenge of ELF. In J. Jenkins, M. J. Dewey, & W. Baker (Eds.), *Routledge handbook of English as a Lingua Franca* (pp. 570–582). Routledge.

Hashimura, *ELT Journal*, 80/1. [full reference to be added]

Iliovits, M., Harding, L., & Pill, J. (2022). Language use in an English-medium instruction university in Lebanon: Implications for the validity of international and local English tests for admissions. *Journal of English-Medium Instruction*, 1(2), 153-179.
<https://doi.org/10.1075/jemi.21009.ili>

Knox, L. (2024, June 18). More Downsizing at Beleaguered ETS. *Inside Higher Ed*.
<https://www.insidehighered.com/news/admissions/graduate/2024/06/18/massive-downsizing-ets-legacy-assessment-company>

Leung, C. (2005). Convivial communication: Recontextualizing communicative competence. *International Journal of Applied Linguistics*, 15(2), 119-144.
<https://doi.org/10.1111/j.1473-4192.2005.00084.x>

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351-372. <https://doi.org/10.1177/026553220101800403>

Harding, L. (2026). Beyond dichotomies: unravelling the complexities of international standardised English language testing. *ELT Journal*, 80(1). Note: Some further edits to this article were made at the production stage.

McNamara, T. (2011). Managing learning: Authority and language assessment.

Language Teaching, 44(4), 500-515. <https://doi.org/10.1017/S0261444811000073>

Nelson, F. (2025, October 28). British Council ‘selling everything it can’ to survive.

Financial Times. <https://www.ft.com/content/af61e4ce-5d10-4300-a84b-1da18dbb2225>

O’Sullivan ... *ELT Journal* 80/1 [full reference to be added]

Shohamy, E. (2001). *The power of tests*. Routledge.

Spolsky, B. (1995). *Measured words: The development of objective language testing*.

Oxford University Press.

Author bio

Luke Harding is a Professor of Applied Linguistics at Lancaster University (UK). His research interests are in language assessment and applied linguistics more broadly, particularly assessing listening and speaking, World Englishes and English as a Lingua Franca, language assessment literacy, and technology and AI in language assessment. Luke served as co-editor of the journal *Language Testing* from 2017-2022 and is co-editor (with Glenn Fulcher) of the Routledge Handbook of Language Testing (2nd Edition). Luke is currently President of the International Language Testing Association (ILTA).