



Looking towards plurilingual futures for literacy assessment

Julie Choi¹ · Russell Cross¹ · Larissa McLean Davies¹ · Sue Ollerhead² · Melissa Barnes³

Received: 27 December 2021 / Accepted: 1 August 2022 / Published online: 13 August 2022
© The Author(s) 2022

Abstract

This paper furthers the understanding of literacy being developed through this Special Issue by bringing a pluralistic, complex account of literacy into conversation with assessment. The rise of neoliberalism in education—and its related focus on standards, accountability, and efficiency—has seen literacy positioned as the “core business” of contemporary education systems and, accordingly, a construct that must be rendered readily measurable, benchmarked, and accounted for. Through a critique of current literacy assessment practices in Australia relying on standardized testing to achieve such accountability using National Assessment Program for Literacy and Numeracy and Literacy and Numeracy Test for Initial Teacher Education as examples, we argue that the resulting dominant account of literacy is both reductive and deficient. On the one hand, we show how standardized literacy testing fails to recognize students’ actual potential, in terms of the diverse and varied plurilingual and cultural meaning-making resources that many students now bring to literacy, beyond using Standard Australian English alone (and as a monolingual speaker would use SAE as an isolated language system). On the other, we show how standardized tests can render the assessment of

✉ Julie Choi
julie.choi@unimelb.edu.au

Russell Cross
r.cross@unimelb.edu.au

Larissa McLean Davies
l.mcleandavies@unimelb.edu.au

Sue Ollerhead
susan.ollerhead@mq.edu.au

Melissa Barnes
melissa.barnes@monash.edu.au

¹ Melbourne Graduate School of Education, University of Melbourne, Melbourne, VIC, Australia

² Macquarie School of Education, Macquarie University, Sydney, NSW, Australia

³ School of Education, Culture & Society, Monash University, Melbourne, VIC, Australia

literacy as effectively meaningless, given difficulties discriminating between test-takers' full range of literacy skills in meaningful ways. We conclude by considering alternative practices with potential to more accurately assess learners' literacy capabilities from a plurilingual perspective, and implications for education policy.

Keywords Literacy assessment practices · Standardized testing · Plurilingual perspectives · Technological affordances

1 Introduction: the rise of standardized testing as a substitute for literacy

This paper furthers the understanding of literacy being developed through this Special Issue by bringing a pluralistic, complex account of literacy into conversation with assessment. The rise of neoliberalism in education—and its related focus on standards, accountability, and efficiency (Corbett, 2008; Polesel et al., 2014) characterized by the Global Educational Reform Movement (GERM) (Sahlberg, 2011)—has seen literacy positioned as the “core business” of contemporary education systems and, accordingly, a construct that must be rendered readily measurable, benchmarked, and accounted for (Parr & Campbell, 2012). Arguments about the apparent decline of school quality are advanced on the basis of data that reveals the reading achievement of the average Australian 15-year-old in 2018 being almost 1 year of schooling behind that of 15-year-old in 2000 (ACER, 2019). Furthermore, the recent PISA 2018 scores show Australian students are below the averages of the highest performing OECD countries such as Singapore, Estonia, Canada, Finland, Ireland, Korea, and Poland—and moreover that 41% of Australian 15-year-olds tested failed to meet the minimum national proficient standards in reading (Thomson et al., 2019).

Statistics like this make for powerful public discourse, yet a global perspective on education begs the question: What does literacy actually *mean* when measured according to standardized assessments? In this article, we show the failure of standardized literacy testing such as National Assessment Program for Literacy and Numeracy (NAPLAN) and Literacy and Numeracy Test for Initial Teacher Education (LANTITE) in Australia to take into account the rich plurilingual, cultural, and multimodal meaning-making resources students now bring to literacy practices (see Lo Bianco in this Issue and Cross, D'warte & Slaughter in this Issue). We also show the inability of such assessments to discriminate between test-takers' full range of literacy skills in meaningful ways. Thus, we argue the resulting dominant account of literacy from such tests is both reductive and deficient. In the final sections, we turn to a discussion of culturally responsive practices to assess more equitably learners' literacy capabilities from a plurilingual perspective.

2 Standardized testing and literacy in schools

In their review of literature on literacy and identity, Moje and Luke (2009) note the significant research that exemplifies the ways in which students' English literacy achievement stands as proxy for the ways in which they and their teachers perceive their identity and academic capacities. Drawing on Hall (1996) and Lewis and del Valle (2009), Moje and Luke (2009, p. 416) observe that,

Because the institutions in which people learn rely so heavily on identities to assign labels of progress, particularly in relation to reading and writing skills, these identity labels associated with certain kinds of literacy practices can be especially powerful in an individual's life.

This phenomenon of equating literacy with student identity, potential, and learning has been amplified exponentially since Moje and Luke's systematic review more than a decade ago, given increasing international reliance on standardized literacy testing as the dominant measure of, and proxy for, student learning and educational achievement. Australia, Canada (Ontario), England, Japan, Scotland, and the USA have all introduced or consolidated their use of standardized, or "census" (McGaw et al., 2020; Perelman, 2018), approaches to literacy assessment during this period, despite there being no correlation between so-called high-performing nations in literacy, and the use of these mandated assessment frameworks (McGaw et al., 2020).

Indeed, research has already established that standardized approaches to testing profoundly disadvantage students who are already vulnerable in the school system. With respect to the Canadian context, both Pinto (2016) and Kearns (2016) report on the introduction of standardized literacy testing as a key accountability measure. Pinto analyzes the negative impacts of the *Ontario Secondary School Literacy Test* (OSSLT) on schools' and teachers' practices over a 20-year period. Similarly, Kearns focuses on the impact of the OSSLT, and considers the ways in which this test affects youth who have failed the test more than once. Arguing that while literacy research and scholarship has emphasized the potential of literacy to free rather than constrain, scholarship in the past three decades has illuminated the importance of multiple literacies beyond print. Kearns demonstrates how the standardized approach of the OSSLT instead renders literacy as colonial, monolingual, and bureaucratic (Kearns, 2016).

In the Australian context, the Federal Government's *Literacy for All* policy (DEETYA, 1998), framed as a response to concerns about apparently declines in literacy performance, advanced six "elements" (p. 10) to improve literacy and numeracy (Henry & Taylor, 1999). These focused on assessment, agreed-upon benchmarks, and reporting against these measures (Cross, 2009, 2011, 2012). Federal funding was subsequently distributed on the basis of "need," determined by having to reach the established benchmarks—paving the way for the introduction of mandated, standardized literacy testing across states. In Victoria, for example, benchmark tests for literacy and numeracy were implemented in 2003 through "Achievement Improvement Monitoring" (AIMs) initiative. In 2008, these and similar state-based tests were replaced by the NAPLAN. NAPLAN was implemented by a federal Labor Government, in the spirit of a policy and practice "revolution" articulating closely with the concerns of the GERM (Sahlberg, 2011), to directly address perceived issues of accountability and transparency in schools (Rudd & Gillard, 2008). Greater surveillance and accountability were further introduced in 2010 when the Government's "myschool" website, which compares the results of students in like contexts, was made publicly available, increasing the impact of the NAPLAN tests on students, teachers, and school communities (Mockler, 2013).

NAPLAN comprises three literacy (writing, reading, and language conventions) and one numeracy test for students in years 3, 5, 7, and 9. Each test is a minimum of 40 min and increases to 65 min for the year 9 test for Reading. Although NAPLAN test results do not impact on a student's progress through school, some jurisdictions, such as New South Wales,

have previously linked achievement in NAPLAN with the award of the post-compulsory Higher School Certificate (Kontominas, 2018). At the time of writing, both New South Wales (Kontominas, 2018) and Victoria (Carey, 2019) intend to schedule an alternative approach to literacy testing in the post-compulsory years to ensure that the literacy requirements of this stage of schooling are being “met,” a move that suggests the curriculum offerings in the senior years have a different understanding of literacy compared to the standardized mechanisms currently used with younger learners.

NAPLAN pre-dated the introduction of the *Australian Curriculum*, the country’s first national curriculum released in draft form in 2010, but now implemented to varying degrees across Australia’s States and Territories (Frawley & McLean Davies, 2015). Given NAPLAN was not aligned with the *Australian Curriculum* until 2017 (McGaw et al., 2020), for most of its first decade, NAPLAN functioned like an intended or official curriculum, being the account of literacy (and numeracy) with which all students and their teachers were engaged and scrutinized against.

Since its inception, NAPLAN has been consistently criticized by the educational community, for its general inefficacy (Lingard, 2009; Thrupp, 2009), the ways in which it drives scorecard-driven behaviors in teachers and schools that limit curriculum offerings (McGaw et al., 2020; Polesel et al., 2014; Thompson & Cook, 2014), and for its impact on student well-being and identity (particularly regarding Indigenous and other students where English is a second or additional language) (Angelo, 2013; Creagh, 2014; Harris et al., 2013). Further, scholars and teachers have offered strong critique of the literacy components of the test, particularly the writing test, which limits responses to the reproduction of formulaic genres (for example Frawley & McLean Davies, 2015; McGaw et al., 2020; Perelman, 2018), diminishing the capacity for students to draw on their own funds of knowledge (González et al., 2005), or animate a rich and generative understanding of twenty-first century literacy (Hipwell & Klenowski, 2011).

The most recent review of NAPLAN (McGaw et al., 2020), initiated by the state governments of New South Wales, Queensland, and Victoria, has recognized many of the complex logistical problems with NAPLAN and the assumptions the tests make, and has made six major recommendations, with related sub-recommendations. The most extensive list of sub-recommendations relates to recommendation 6, concerning NAPLAN’s writing test, which McGaw’s panel suggests be fully redeveloped. However, although the recommendations set out in the Review do address several major issues, including choice of genres, the time allowed for writing, curriculum alignment, and the use of digital technologies (p. 144), they do not extend to fundamental issues on the ways in which NAPLAN marginalizes certain students, rendering non-standard, non-dominant forms of literacy knowledges, experiences, and skills irrelevant to schooling and progress. Indigenous students, for example, are only mentioned once within the recommendations, and then only the context of continuing to publish performance levels differentiated by cohort profiles. Similarly, Schalley et al. (2015, p. 168) present a powerful argument for how the limitations of “NAPLAN are aggravated when it comes to minority language speaking children,” not only contributing to a narrowing of literacy that ignores multilingualism as an asset, but results in data informed by an “impoverished approach to literacy education” (p. 169).

Thus, even with the modifications suggested, it is difficult to see how NAPLAN tests will not continue to show educators what they already know; namely, that the test is limited in being able to recognize the full breadth of individual ability, that some students are likely to

achieve higher (or lower) scores based on assumptions about background and context, and that definitions of literacy in the Australian education will continue to be dominated by how it is conceived as a focus of high-stakes testing taken periodically by students through the years of schooling.

3 Problems of standardized literacy testing: LANTITE and teachers

In addition to the impact of standardized testing on literacy education in schools outlined above in the case of NAPLAN, examples of its limitations also extend to higher education—particularly as it relates to the testing of prospective *teachers'* literacy. Standardized tests of basic literacy and/or numeracy have been used as entry hurdles to the teaching profession in many countries since the early 2000s, including the US Pre-professional Skills test (Praxis I) (Gitomer et al., 2011; Tatto, 2015), and UK Qualified Teacher Status Numeracy and Literacy Test (McNamara et al., 2002). The UK Department for Education rescinded the requirement that prospective teachers complete the Qualified Teacher Status Professional Skills Test from April 2020, however, with the onus instead being on teacher education providers to “assure a candidate’s fundamental English and mathematics either before or during their course” (United Kingdom Department of Education, 2021, para. 1). Similarly, China has also announced its intention to remove its teacher qualification test for graduates from endorsed programs of teacher education, although all civil servants, including teachers, are still required to undertake the National Mandarin Language Test and reach a certain level to pass, based on their teaching area specialism (CIEB, 2020; Xinhua, 2020).

In this section, we build on the earlier discussion with respect to NAPLAN, but with reference to the Australian LANTITE as a case to illustrate further limitations of standardized literacy testing in the context of professional gatekeeping hurdles (see also Barnes & Cross, 2022). LANTITE was introduced in 2016 by the federal Australian government as a mechanism to determine whether those entering the teaching profession “have the required personal literacy and numeracy levels to work as a teacher in Australia” (VIT, 2021). One in a series of reform initiatives developed in response to a Commonwealth ministerial review of Australian teacher education, LANTITE aims to promote greater transparency for how students gain entry into programs of teacher education (TEMAG, 2014).

As alluded to earlier in this paper, the years preceding LANTITE’s introduction saw intense public and media scrutiny on falling national and international rankings of educational competitiveness and school performance (Baroutsis & Lingard, 2017; Gorur & Wu, 2015; Robinson, 2018a). This commentary was accompanied by a broader narrative of failure around perceived problems of inept teachers and teacher quality, alongside skepticism of whether university-based teacher education—critiqued for their focus on “nebulous gobbledygook” and edu-babble—was ensuring its graduates were “classroom ready” (Gale & Cross, 2007; Gale & Parker, 2017).

In contrast to other professional disciplines, such as law and medicine, with a long history of using standardized tests to competitively select the highest-performing students from a large pool of applicants vying for the few positions available in each round, teacher education has been heavily criticized for being positioned as the university ‘cash cow’ (Dinham, 2013; Zyngier, 2016), intentionally oversubscribing student enrolments into comparatively cheap, easy-to-fill course places for financial gain. This has resulted in a number of teacher education programs modifying their entry standards (Dinham, 2013), with some students having gained admission

despite having not met standard entry requirements (Robinson, 2018b). LANTITE was therefore developed to better regulate who entered the teaching profession at the point of teacher preparation, with a focus on ensuring only “the best” candidates would go through to become teachers (TEMAG, 2014, p. x). The “best,” as defined by LANTITE on the basis of the TEMAG report, and of specific relevance to the issues in this paper, are those with “levels of personal literacy and numeracy ... broadly equivalent to those of the top 30 per cent of the population” (AGDET, 2016, p. 6).

LANTITE consists of 14 test templates constructed from a bank of 108 test items, with 60 items addressing “literacy”—40 on “reading” and 20 on “writing” (ACER, 2021a). Full details on LANTITE’s structure, design, and sample test items can be found at teacheredtest.acer.edu.au. Of note, the test does not directly assess applicants’ writing (i.e., no writing samples are submitted or examined) but it instead assesses the applicants’ knowledge of the technical components and structures of writing through multiple-choice questions.

There are many well-established critiques in the literature cautioning against the use of standardized tests as a gatekeeping measure, especially in the field of teaching, including evidence that minorities tend to be over-represented in groups most adversely affected by these gatekeeping measures (Goldhaber & Hansen, 2010; Graham, 2013; Petchauer & Baker-Doyle, 2016; van Gelderen, 2017). However, our primary critique is how *ineffectual* LANTITE seems to be as a tool to offer any useful discriminatory capacity of candidates’ literacy skills. Of the 5000 candidates in the pilot cohort, for example, 92% passed the literacy component, which rose to an even higher near 95% pass rate following its official implementation (Barry, 2017; Knott, 2016, para. 5).

Previous research detailing a quantitative analysis of 2013 LANTITE test results from students at a large metropolitan university in Melbourne, Australia (Barnes & Cross, 2020), found the overall proportion of students who failed the test (5–10%) aligned with the national average (Barry, 2017; Knott, 2016). However, it also examined the effect of including subsequent re-sit outcomes on the overall pass rates. That is, although the 9.9% fail rate for literacy at the institution fell within the 5–10% national average (195 students out of 2103), this was only the outcome for the students’ first attempt. Teacher candidates who fail are then allowed, by default, three additional attempts to pass the LANTITE; with a letter of support from their institution, they can apply to sit the test up to five times (ACER, 2021b).

The analysis in Barnes and Cross (2020) revealed that students had a 50% chance of passing the test on each subsequent attempt: at the time of data collection, 31 of the original 195 students who failed the literacy test re-sat the test, with 17 of those passing the test on their second attempt. Moreover, of the 14 students who failed that second re-sit, 7 re-sat the test for a third attempt and 4 passed the test. This left only 3 of the original 31 who attempted re-sits to fail all three re-sits. At the time the data was being collected, one of the three had received approval to take the test a fourth time and was preparing to do so. Statistically, the 9.9% fail rate for literacy narrows closer to 5% after students have exhausted their opportunities to re-sit the test.

With a pass rate of 95%, questions arise about the adequacy of LANTITE as a valid gatekeeping measure for literacy. In addition to its negligible impact on being able to effect substantive reform, given essentially the same number of applicants go on to gain entry into teacher education despite the test, many candidates hold concerns about the test’s validity. Barnes and Cross (2020, p. 316) report one teacher candidate who sees the test as valid given its real-world application for teachers, with respect authenticity. However, others countered

that the test seemed heavily dependent on one's prior experiences, and being exposed to certain texts at school rather than others (p. 316):

I didn't realise when I sat it that it would all be in context of what you actually do as a teacher. Like reading NAPLAN results off a graph. We can't deny that those skills are important.

If you're not someone who is an avid reader, you've never seen that word before in your life, and it's a disadvantage. It's not really showing their literacy skills, it just means that they haven't been exposed to that certain word and then they can't answer the question because it's asking for a synonym of it.

For these reasons, a number of teacher candidates felt that being from a multilingual background, or completing their schooling outside of Australia where they may not have been exposed to the same literacy texts, has been a disadvantage (Barnes & Cross, 2020). LANTITE's overemphasis on vocabulary, for example, was deemed unfair for those without exposure to the same vocabulary prioritized in Australian schooling contexts, yet a heavy focus on language elements is how literacy has been conceptualized in the test. In several focus groups, Barnes and Cross (2020) report teacher candidates arguing that studying in another country and/or having a Language Background Other than English background could put students at a disadvantage because, even if these students were still able to pass the test, it required additional time to study specifically for the test and its language focus (even though it claims to be a test of literacy), as well as need for potential multiple re-sits. These concerns echo those of scholars around the world that find persistent discrimination of teachers and graduate teachers from English second or additional language backgrounds, and the additional hurdles they continually confronted with during their professional careers due to being non-native speakers of English (e.g., Cho, 2010 ; Hélot, 2017 ; Tupas, 2015).

A standardized literacy test such as LANTITE that consists entirely of multiple-choice questions with an emphasis on language is ultimately limited in what it can assess and therefore has a narrowed conception of what constitutes "literacy" (Barnes & Cross, 2018). As already highlighted earlier, there is also significant evidence that multilingual speakers are often discriminated against in regard to their English language competence (Cho, 2010; Hélot, 2017; Tupas, 2015), in contrast to having this competence valued as an asset: additional professional expertise. The limitations of standardized tests—and their narrow focus on what can be readily tested as "literacy"—compounds these concerns. Even if multilingual speakers themselves are not directly disadvantaged by such tests, given almost all applicants can still pass, as a gatekeeping solution, it sends strong messages about "what skills" are the most valued for the profession, and who is the "best kind of applicant": in this case, a monolingual (and therefore "native" speaker) account of literacy and literacy skills.

Many of the teacher candidates have argued, irrespective of their own native/non-native speaker status, that LANTITE impacted how they viewed themselves and the profession (Barnes, 2021; Barnes & Cross, 2020). They lamented the importance placed on basic skills that were measured through a high-stakes test that almost everyone can pass, while also having to find the resources—both time and financial—to satisfy the hurdle. Ultimately, it seemed a costly reminder of the distrust of their teacher competence. As one teacher candidate argued, LANTITE seemed to "set the teaching profession back" and, because of the media attention given to the need for this test, it had positioned them as incompetent: "They [the general

public, policymakers, etc.] basically don't think you're literate" (Barnes & Cross, 2020, p. 318). Another reiterated: "It's just the government being...people being, like oh, teachers don't know how to read and write" (p. 318).

Ultimately, as with NAPLAN, it is important to consider whether the benefits of the policy outweigh their unintended consequences. While we believe there is a need for the best people in the teaching profession, a standardized test such as LANTITE might not be the *means* of achieving this aim. Broader concerns about whether the test even achieves substantial reform aside (Barnes & Cross, 2022), we argue that the standardized literacy testing of teachers with respect to multilingual speakers results in a blindspot that fails to recognize applicants with broader linguistic and cultural profiles that might be of greater benefit to certain learners and communities (including monolingual-centric communities without multilingual teacher role models, in terms of aspiration).

4 Why the need for a plurilingual perspective on literacy assessment?

Given the issues outlined above, we now turn to what an alternate "plurilingual" perspective on assessment might offer, drawing on international studies that addresses issues of difference and diversity in school level literacy assessments. As argued throughout this Issue, as globalization and immigration increases, the ability to communicate in multiple languages has become an increasing feature of school settings. Yet current paradigms for assessing the language and literacy competence of multilingual learners, such as those informing NAPLAN and LANTITE, lag behind the most recent views of what it means to learn a second language, and what it means to know multiple languages (Canagarajah, 2006; Shohamy, 2013), ignoring the reality of how languages are being used by multilingual speakers around the world in the context of migration and globalization (Shohamy, 2011). A failure to acknowledge the critical role that language proficiency plays in poor exam performance resonates across contexts where English is the dominant language of assessment (Menken, 2008). As studies in tests on language policy show, "policymakers use tests to create *de facto* policies that will promote their agendas and communicate their priorities, a top-down practice which Shohamy (Shohamy, 1998, 2001) characterizes as unethical, undemocratic, and unbeneficial to the test-taker" (Menken, Menken, 2008, p. 404).

The dominant approaches to literacy assessment are based on a monoglossic view of language; that is, language(s) understood as being separate entities (i.e., "English" in these examples of NAPLAN and LANTITE, or more specifically, "Standard Australian English"), ignoring the complex communicative practices of multilinguals and their simultaneous uses of multiple languages (Shohamy, 2013). A plurilingual perspective on literacy, as again outlined in other articles within this Issue, rejects monocentric, monoglossic views of language, and language use, which idealize communicative native speakers' norms. Plurilingualism, understood as the unitary linguistic and sociolinguistic ability of individuals to use more than one language in everyday and academic contexts, positions learners' lived experiences, interests, knowledges, capabilities, and their own ways of making meaning at its core. Educators therefore similarly position learners as not only multilingual but "multicompetent": "an individual with knowledge of an extended and integrated linguistic repertoire who is able to use the linguistic variety for the appropriate occasion" (Franceschini, 2011, p. 351). Pushing against monolingual curriculum structures, educators create classroom conditions where multilingual learners are encouraged and supported to deploy all their available knowledge

resources through challenging learning tasks. There are now many studies on the benefits of such pedagogies for cognitive and identity development for multilingual learners in various school settings and higher education contexts (see Cenoz & Gorter, 2011; Creese & Blackledge, 2015; Cummins, 2017).

At the level of assessment, the understanding of enacting learners' whole range of meaning making resources is often still unrecognized and forbidden in de facto monolingual tests, especially standardized testing, as discussed earlier. Despite ontological understandings of bilinguals as having "a specific linguistic configuration characterized by the constant interaction and coexistence of the two languages involved" (Herdina & Jessner, 2002, p. 59), standardized tests such as NAPLAN as well as LANTITE of literacy that dominate Australian educational discourse continue to see the presence of other languages as "irrelevant to their capacity to perform in the target language" (Schissel et al., 2018, p. 177). Furthermore, others argue that testing multilingual students in a language that they are still learning, against the same criteria as their native speaker of English counterparts, is to "compare groups of incomparable conditions" (Shohamy, 2011, p. 419). Put simply, *monolingual tests* offer invalid evidence of what *multilingual students* know and can do.

5 Implications for culturally responsive assessment practices: assessing children's multilingual competency

Assessing children's multilingual competency in their home language(s) alongside the dominant language of schooling has significant implications for cultural identity and well-being. There is consensus that children who speak multiple languages benefit from enhanced cognitive skills (e.g., executive functioning and working memory) and social skills, including enhanced relationships with family and community members (De Houwer, 2015; Puig, 2010).

As Gorter and Cenoz (2017) observe with respect to the implications of this for assessment, tests should align with multilingual children's actual language *practices*, rather than an artificial construct of monolingual language use. Assessment for multilingual children should acknowledge and reference how children draw upon their plurilingual repertoires, such as communicating across more than one named language to understand classroom content or complete a project. In such cases, the assessment of (plurilingual) competence has (contextual) validity.

Seed (2019) suggests that criteria can focus on an individual's plurilingual abilities to demonstrate degrees of proficiency in a standard "named" language. For example, if a learner intersperses home language vocabulary within longer stretches of the standard, dominant language discourse, but the overall sense of the response is correct, that would offer an indication (and recognition of) the learner having exercised *plurilingual competence*, to communicate in the standard language—rather than simply on what was done in the target language alone as the sole indicator of competence, or ability, as the measure of attainment.

Despite a plethora of research literature on the "multilingual turn in education" (Conteh & Meier, 2014; May, 2014), to date, there has been disproportionately less consideration given to plurilingual assessment practices. It is therefore nigh impossible to present a review of empirical studies on the effectiveness of different approaches to plurilingual assessments, or to draw concrete implications for practice. Yet as researchers and educators grapple with the unacceptable use of artificially monolingual test practices to measure more authentic,

plurilingual instances of language use, a growing number of assessments in plurilingual contexts have begun to emerge, which illustrate the importance of differentiating different types of assessment for gauging different instances of language learning and skills development. In a 2020 edition of Cambridge Research Notes titled “What does plurilingualism mean for language assessment?,” Seed presents a framework that groups plurilingual assessments into four main categories, which will be briefly summarized below.

The first category assesses the ability to use one’s plurilingual repertoire to “aid learning or proving skills in *one named language*” (Seed, 2020, p. 9). An example would be an English assessment where a French speaker with beginner-level English proficiency draws on their French knowledge to produce the phrase “My friends are very *sympathetic*,” to mean “My friends are very *nice*.” Instead of seeing the test-taker’s use of the word “sympathetic,” a false friend to the French word “*sympa*” (meaning “nice”), as an error, this utterance can instead be regarded as evidence of the test-taker’s productive use of plurilingual knowledge in an attempt to communicate in English.

The next category Seed recognizes is that of “drawing on one’s plurilingual repertoires to aid learning or proving skills in *more than one named language*” (p. 10). Assessments in this category can be used in plurilingual contexts where the use of two or more languages is the norm, and where it is appropriate to use what the Common European Framework of Reference (Council of Europe, 2020) terms “cross-linguistic mediation.” This could involve the translation of a simple text from one language into another, summarizing in one language the main points of a conversation conducted in another language or writing in one language the relevant points in an article in a different language.

The third category of assessments involves those that measure content learning outcomes, rather than language learning or skills, where the instructions and questions are written in two or more languages and test takers can choose which language they use to answer. Such types of assessment practices are primarily concerned with fairness, so that test takers are not prevented from demonstrating their knowledge of subject content as a result of not being proficient in the dominant language of instruction. An example can be seen in the multilingual large-scale assessments of students’ knowledge in three languages in South Africa, a superdiverse context with eleven official languages (Heugh et al., 2017).

The final type of plurilingual assessment that Seed (Seed, 2020) outlines is the development of “plurilingual competence to function with languages not known, or only partially known.” This type of assessment differs from the three mentioned previously, in that it refers to attitudes and behaviors towards building up general communicative competence and self-efficacy in learning additional languages. This may include a positive attitude towards exposing oneself to new languages, as well as strategies such as using cognates and internationalisms to make meaning in an unfamiliar language. Assessment techniques in this category would include qualitative methods such as observations, monitoring, learning portfolios, learning diaries, and self-assessment (p. 12).

An assessment of each language that a learner speaks may show different strengths in particular domains distributed unevenly across languages (Kohnert & Bates, 2002). Distributed skills and uneven ability across the different languages a multilingual learner speaks have significant implications for the ways in which we assess them. We need to regularly assess all languages (not just the language of instruction) to gain an accurate picture of language capacities. Australian schools’ tendency to provide high stakes, once-off snapshots of multilingual children’s English literacy ability (such as NAPLAN in years 3, 5, 7, and 9) and to base pedagogical decisions based on this data contradicts this research advice.

Due to this uneven distribution of language development across multiple languages, it is essential that when assessing a multilingual learner, a thorough and culturally sensitive case history is conducted (Shiple & McAfee, 2020). This should include the age at which a child was exposed to each language, the amount of exposure to and use of language on a typical day, the people who speak each language to the child (parents, siblings, carers, friends), the settings or contexts for language use (e.g., home, religious settings, community groups, or school), and the child's preferred language (e.g., for home, religious settings, community groups, or school).

6 Technological affordances and literacy assessment for plurilingual learners

The above highlights the exceedingly individual and situated nature of plurilingual learners' literacy in ways that suggest that assessment must also be personalized to each learner, and localized to contexts where plurilingual practices are demonstrated in action (Seed, 2019). Although operationalizing such constructs into a test or attempting to measure diversity in a standard way appears oxymoronic, García and Flores (2013, p. 162) argue that possibilities do exist given the rapid and sophisticated development of what new technologies can offer.

Technology is assisting with the development of Internet-based adaptive tests that can adapt the language load, as well as the language use, to the bilingual students' linguistic profile. Thus, the language of the test can be simplified, translated, or changed to adjust to the students' languaging to ensure the assessment of language and literacy use, rather than just discrete language skills, and the assessment of content proficiency and knowledge independent of language, Internet-based adaptive tests are capable of being flexible multilingual tests that adjust to students' dynamic language practices. These tests can also provide visuals and glossaries to contextualize language for bilingual students.

Highlighting the importance of testing "the ability-in language user-in context" orientation, that is, the entanglement or inseparability of a person's ability in interaction with a particular task within a particular context, scholars are also turning their attention to the possibilities of game-based assessment (GBA) as a method of enacting "dynamic, tailored, and multisensory experience" (Lay et al., 2017). Games can be designed in context-rich ways using "avatar creation, complex scenarios, and extended play" that draw out the "intricacies of a person by task interaction within complex and dynamic digital environments" (p. 9; see also Chiu et al.'s (2012) study of a meta-analysis into English as a Foreign Language game-based learning in Asia). While a handful of studies (e.g., DiCerbo, 2014; DiCerbo et al., 2016; Mislevy et al., 2012; Mislevy et al., 2015) show the rich possibilities of technology-based dynamic assessments, as Lay et al. (2017) state,

A survey of the published literature uncovers practically no publications focusing on L2 GBA. This is perhaps not surprising given the recent engagement in this domain, including in L2 learning—and testing typically follows versus leads in exploring and adopting innovations ... Nonetheless, explorations of the scant GBA literature available in the wider measurement literature can be informative. (p. 10)

Although resource is intensive in terms of investment and development, these advances nonetheless provide evidence of the potential for technology to address the challenges identified for assessing plurilingual students' competency in literacy, at scale. There is a need for a commitment to collaborating with experts in the relevant fields of studies—language, assessment, and

technology—otherwise, we risk remaining with a reliance on easy-to-use but ultimately blunt instruments that reveal little about the actual competence of learners now populating our education system. Shohamy and Pennycook (2019) state, “testing has only recently started to address this question of testing repertoires rather than languages” (p. 33) and multilingual language tests are “somewhat more complex to design and administer” (p. 35). Thus, we are still light years away from experiencing such orientations at the level of national assessments, but so far the “results show that they are more equitable tests of multilingual abilities” (p. 35).

7 Conclusion

Although curriculum and teaching have begun, albeit slowly, to engage with approaches that better recognize and embrace learners’ different and diverse ways of making meaning (see Beavis, Bacalja & O’Brien in this Issue), assessment practices are still far behind, dominated by tools that advance standardized but reductive forms of literacy skills and competence. Yet, as Gorter and Cenoz (2017) argue, “tests should match actual language practices and multilinguals use resources from their whole linguistic repertoire. If teaching is going in the direction of a multilingual focus, assessment should also follow the same path” (p. 243).

Alternative knowledge and technologies are now available to address this challenge, but it is not yet clear what these tests should look like. We suggest that turning to insights from plurilingual perspectives is an important first step. As Lopez et al. (2017) argues, for example, this includes acknowledging learners for their ability to negotiate relationships between a variety of different languages, while Shohamy (2011) argues of the need for any form of assessment to match the language varieties that learners currently use rather than a native speaker standard that is neither achievable nor sought after. As Mathew (Mathew, 2008, pp. 32–33) summarizes,

An assessment system that has as its guiding principles open and flexible language, interactiveness, multiple approaches, dynamic assessment, and integration of assessment with instruction is recommended. In sum then, assessment has to adopt a multiple approach: multi-task, multi-rater, multi-candidate and multi-norms. We need to move from the ‘either/or’ orientation to a ‘both and more’ perspective. We need to shift the emphases from language as system to language as social practice, from grammar to pragmatics, from competence to performance in our attitude to proficiency.

This paper serves to emphasize the complex, multi-dimensional nature of literacy beyond the compartmentalization of single “languages,” like those promulgated through current mainstream assessments that dominate Australian discourse on literacy. Such tests, we argue, are rendered unfit for purpose for our plurilingual society, compromising their significance and potential. In an educational context that emphasizes values of diversity, inclusion, and equity, developing alternative, more equitable and valid forms of literacy assessment for our plurilingual learners is both timely and urgent.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AGDET. (2016). *Quality schools, quality outcomes*. Retrieved from: https://docs.education.gov.au/system/files/doc/other/quality_schools_acc.pdf
- Angelo, D. (2013). NAPLAN implementation: Implications for classroom learning and teaching, with recommendations for improvement. *TESOL in Context*, 23(1/2), 53–73.
- Australian Council for Education Research [ACER]. (2021a). *Literacy and numeracy test for initial teacher education students*. Retrieved from <https://teacheredtest.acer.edu.au>
- Australian Council for Education Research [ACER]. (2021b). *Re-sit*. Retrieved from <https://teacheredtest.acer.edu.au/results/re-sit>
- Australian Council for Education Research [ACER]. (2019). *PISA 2018: Australian students' performance*. Retrieved from <https://www.acer.org/au/discover/article/pisa-2018-australian-students-performance>
- Barnes, M. (2021). Policy actors or objects of policy? Teacher candidates' interpretations of 'teacher quality' policy initiatives in Australia. *Teaching and Teacher Education*, 106. <https://doi.org/10.1016/j.tate.2021.103440>
- Barnes, M., & Cross, R. (2018). *Why we need to review how we test for teacher quality*. The conversation. <https://theconversation.com/why-we-need-to-review-how-we-test-for-teacher-quality-95074>
- Barnes, M., & Cross, R. (2020). Teacher education policy to improve teacher quality: Substantive reform or just another hurdle? *Teachers and Teaching*, 26(3–4), 307–325.
- Barnes, M., & Cross, R. (2022). Standardised testing as a gatekeeping mechanism for teacher quality. In I. Menter et al. (Eds.), *The Palgrave handbook of teacher education research* (pp. 1–18). Palgrave.
- Baroutsis, A., & Lingard, B. (2017). Counting and comparing school performance: An analysis of media coverage of PISA in Australia, 2000–2014. *Journal of Education Policy*, 31(4), 432–449.
- Barry, S. (2017). *New teachers score 95 percent in skills test*. Retrieved from <https://www.school-news.com.au/news/new-teachers-score-95-percent-in-skills-test/>
- Canagarajah, S. (2006). Toward a writing pedagogy of shuttling between languages: Learning from multilingual writers. *College English*, 68(6), 589–604.
- Carey, A. (2019). *Students set to face new literacy and numeracy tests in VCE revamp*. Retrieved from <https://www.theage.com.au/national/victoria/students-set-to-face-new-literacy-and-numeracy-tests-in-vce-revamp-20190909-p52pm2.html>
- Cenoz, J., & Gorter, D. (2011). A holistic approach to multilingual education. Introduction. *The Modern Language Journal*, 95(3), 339–343.
- Center on International Education Benchmarking [CIEB]. (2020). *Shanghai-China: Teacher and principal quality*. Retrieved from <https://ncee.org/center-on-international-education-benchmarking/top-performing-countries/shanghai-china/shanghai-china-teacher-and-principal-quality/>
- Cho, C. (2010). "Qualifying" as teacher: Immigrant teacher candidates' counter-stories. *Canadian Journal of Educational Administration and Policy*, 100(3), 1–22.
- Chiu, Y. H., Kao, C. W., & Reynolds, B. L. (2012). The relative effectiveness of digital games-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology*, 43, 104–107.
- Conteh, J., & Meier, G. (Eds.) (2014). The multilingual turn in languages education: Opportunities and challenges. Multilingual Matters.
- Corbett, M. (2008). The Edumometer: The commodification of learning from Galton to the PISA. *Journal for Critical Education Policy Studies*, 6(1), 352–378.
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. .
- Creagh, S. (2014). A critical analysis of problems with the LBOTE category on the NAPLAN test. *The Australian Educational Researcher*, 41(1), 1–23. <https://doi.org/10.1007/s13384-013-0095-y>

- Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics*, 35, 20–35.
- Cross, R. (2009). *Literacy for all: Quality language education for few*. *Language and Education*, 23(6), 509–522.
- Cross, R. (2011). Monolingual curriculum frameworks, multilingual literacy development: ESL teachers' beliefs. *Australian Journal of Language & Literacy*, 34(2), 166–180.
- Cross, R. (2012). Reclaiming the territory: Understanding the specialist knowledge of ESL education for literacy, curriculum, and multilingual learners. *TESOL in Context*, 22(1), 4–17.
- Cummins, J. (2017). Teaching minoritized students. Are additive approaches legitimate? *Harvard Educational Review*, 87(3), 404–425.
- De Houwer, A. (2015). Harmonious bilingual development: Young families' well-being in language contact situations. *International Journal of Bilingualism*, 19(2), 169–184.
- Department of Employment, Education, Training and Youth Affairs [DDETYA]. (1998). Literacy for all. Australian Commonwealth Government.
- Dinham, S. (2013). *Standards will slide while teacher education is used as a cash cow*. Retrieved from <https://theconversation.com/standards-will-slide-while-teacher-education-is-used-as-a-cash-cow-11677>
- DiCerbo, K. E. (2014). Game-based assessment of persistence. *Educational Technology & Society*, 17(1), 17–28.
- DiCerbo, K. E., Mislevy, R. J., & Behrens, J. T. (2016). Inference in game-based assessment. In H. F. O'Neil, E. L. Baker, & R.S. Perez (Eds.), *Using games and simulations for teaching and assessment* (pp. 253–279). Routledge.
- Franceschini, R. (2011). Multilingualism and multicompetence: A conceptual view. *The Modern Language Journal*, 95(3), 344–355.
- Frawley, E., & McLean Davies, L. (2015). Assessing the field: Students and teachers of writing in high-stakes literacy testing in Australia. *English Teaching: Practice & Critique*, 14(2), 83–99. <https://doi.org/10.1108/ETPC-01-2015-0001>
- Gale, T., & Cross, R. (2007). Nebulous goobledgeook: The politics of (re)learning how and what to teach. A. Berry, A. Clemans & A. Kostogriz (Eds.), *Dimensions of professional learning: Professionalism, practice and identity* (pp. 5-21). Sense Publishers.
- Gale, T., & Parker, S. (2017). Retaining students in Australian higher education: Cultural capital, field distinction. *European Educational Research Journal*, 16(1), 80–96. <https://doi.org/10.1177/1474904116678004>
- García, O., & Flores, N. (2013). Multilingualism and common core state standards in the United States. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL, and bilingual education* (pp. 147–166). Routledge.
- Gitomer, D., Brown, T., & Bonett, J. (2011). Useful signal or unnecessary hurdle? The role of basic skills tests in teacher preparation. *Journal of Teacher Education*, 62, 431–445.
- Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing. *American Educational Research Journal*, 47(1), 218–251.
- González, N., Moll, L., & Amanti, C. (2005). Funds of knowledge. Erlbaum Associates.
- Gorter, D., & Cenoz, J. (2017). Language education policy and multilingual assessment. *Language and Education*, 31(3), 231–248.
- Gorur, R., & Wu, M. (2015). Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse*, 36(5), 647–664.
- Graham, A. (2013). Black teacher education candidates' performance on praxis I. *The Negro Educational Review*, 64(1), 9–35.
- Hall, S. (1996). Who needs "identity"? In S. Hall & P. D. Gay (Eds.), *Questions of cultural identity* (pp. 1–17). Sage.
- Harris, P., Chinnappan, M., Castleton, G., Carter, J., De Courcy, M., & Barnett, J. (2013). Impact and consequence of Australia's National Assessment Program – Literacy and Numeracy (NAPLAN) – using research evidence to inform improvement. *TESOL in Context*, 23(1/2), 30–52.
- Henry, M., & Taylor, S. (1999). From social justice to 'literacy for all': Australian variations on a global theme. Retrieved from <https://www.aare.edu.au/data/publications/1999/tay99032.pdf>
- Herdina, P., & Jessner, U. (2002). A dynamic model of multilingualism: Perspectives of change in psycholinguistics. *Multilingual Matters*.
- Hélot, C. (2017). Awareness raising and multilingualism in primary education. In J. Cenoz, D. Gorter & S. May. (Eds.), *Language awareness and multilingualism. Encyclopedia of Language and Education* (3rd ed., pp. 247–262). Springer.
- Heugh, K., Prinsloo, C., Makgamatha, M., Diedericks, G., & Winnaar, L. (2017). Multilingualism(s) and system-wide assessment a southern perspective. *Language and Education*, 31(3), 197–216.
- Hipwell, P., & Klenowski, V. (2011). A case for addressing the literacy demands of student assessment. *Australian Journal of Language and Literacy*, 34(2), 127–146.

- Kearns, L. (2016). The construction of ‘illiterate’ and ‘literate’ youth: The effects of high-stakes standardized literacy testing. *Race Ethnicity and Education*, 19(1), 121–140. <https://doi.org/10.1080/13613324.2013.843520>
- Knott, M. (2016). ‘A total shambles’: New literacy and numeracy test for teachers frustrate students, educators. The Sydney Morning Herald. Retrieved from <https://www.smh.com.au/politics/federal/a-total-shambles-new-literacy-and-numeracy-test-for-teachers-frustrates-students-educators-20160901-gr6hi7.html>
- Kohnert, K., & Bates, E. (2002). Balancing bilinguals: Lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech Language and Hearing Research*, 45(2), 347–359.
- Kontominas, B. (2018) NAPLAN no longer linked to HSC after backlash from parents. ABC news. Retrieved from <https://www.abc.net.au/news/2018-02-22/naplan-no-longer-linked-to-hsc/9473238>
- Lay, A., Paton, E., & Chelhouh-Deville, M. (2017). A case for the use of the ability-in language user-in context orientation in game-based assessment. *Language Testing in Asia*, 7(16). <https://doi.org/10.1186/s40468-017-0045-0>
- Lingard, B. (2009, November). Testing times: The need for new intelligent accountabilities for schooling. *Queensland Teachers Professional Magazine*, 24, 13–19.
- Lewis, C., & del Valle, A. (2009). Literacy and identity: Implications for research and practice. In L. Christenbury, R. Bomer & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 307–322). Guilford Press.
- Lopez, A.A., Turkan, S., & Guzman-Orth, D. (2017). Assessing multilingual competence. In E. Shohamy, I. Or, & S. May (2016). *Language testing and assessment. Encyclopedia of language and education* (3rd ed., Pp. 1–11). Springer International Publishing AG.
- Mathew, R. (2008). Assessment in multilingual societies. In E. Shohamy & N.H. Hornberger (Eds.), *Encyclopedia of language and education, Volume 7: Language testing and assessment* (2nd ed., pp. 19–36). Springer Science+Business Media LLC.
- May, S. (Ed.) (2014). *The multilingual turn: Implications for SLA, TESOL and Bilingual education*. Routledge.
- McGaw, B., Louden, W., & Wyatt-Smith, C. (2020). *NAPLAN Review Final Report*. Retrieved from https://naplanreview.com.au/pdfs/2020_NAPLAN_review_final_report.pdf
- McNamara, O., Roberts, R., Basit, T., & Brown, T. (2002). Rites of passage in initial teacher training. *British Educational Research Journal*, 8(6), 863–878.
- Menken, K. (2008). High-stakes tests as de facto language education policies. In E. Shohamy & N.H. Hornberger (Eds.), *Encyclopedia of language and education, Volume 7: Language testing and assessment* (2nd ed., pp. 401–413). Springer Science+Business Media LLC.
- Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., Frezzo, D. C., & West, P. (2012). Three things game designers need to know about assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 59–81). Springer.
- Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A., & John, M. (2015). Psychometrics and game-based assessment. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 23–48). Routledge.
- Mockler, N. (2013). Reporting the ‘education revolution’: MySchool.edu.au in the print media. *Discourse: Studies in the Cultural Politics of Education*, 34(1), 1–16. <https://doi.org/10.1080/01596306.2012.698860>
- Moje, E., & Luke, A. (2009). Literacy and identity: Examining the metaphors in history and contemporary research. *Reading Research Quarterly*, 44(4), 415–437.
- Parr, M., & Campbell, T. (2012). Understanding literacy as our WORLD inheritance: Re-visioning literacy discourse and its implications for teaching practice. *International Review of Education*, 58(4), 557–574.
- Perelman, L. (2018). Towards a new NAPLAN: Testing to the teaching. *NSW Teachers Federation*.
- Petchauer, E., & Baker-Doyle, K. (2016). Next thing you know, her hair turned green. *Critical Studies in Education*, 1–18. <https://doi.org/10.1080/17508487.2016.1234496>
- Pinto, L. (2016). Tensions and fissures: The politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, 27(1), 95–112. <https://doi.org/10.1080/09585176.2016.1140061>
- Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640–657. <https://doi.org/10.1080/02680939.2013.865082>
- Puig, V. I. (2010). Are early intervention services placing home languages and cultures “at risk”? *Early Childhood Research and Practice*, 12(1) Retrieved from <http://ecrp.uiuc.edu/v12n1/puig.html>
- Robinson, N. (2018a). Calls for NAPLAN review after report reveals no change in decade of results. Australian Broadcasting Company. Retrieved from <https://www.abc.net.au/news/2018-03-07/naplan-call-review-after-report-reveals-no-change-in-decade/9519840>
- Robinson, N. (2018b). Students with lowest ATAR scores being offered places in teaching degrees: Secret report. Australian Broadcasting Company. Retrieved from <https://www.abc.net.au/news/2018-09-18/students-lowest-atar-scores-teaching-degree-offers-secret-report/10200666>

- Rudd, K., & Gillard, J. (2008). Quality education: The case for an education revolution in our schools. Commonwealth of Australia.
- Sahlberg, P. (2011). Finnish lessons: What can the world learn from educational change in *Finland*? Teachers' College Press.
- Schalley, A. C., Guillemin, D., & Eisenclas, S. A. (2015). Multilingualism and assimilationism in Australia's literacy-related educational policies. *International Journal of Multilingualism*, 12(2), 162–177. <https://doi.org/10.1080/14790718.2015.1009372>
- Schissel, J., Leung, C., López-Gopar, M., & Davis, J. R. (2018). Multilingual learners in language assessment design for linguistically diverse communities. *Language and Education*, 32(2), 167–182.
- Seed, G. (2019). *What is plurilingual assessment? Paper presented at the Association of Language Testers in Europe*.
- Seed, G. (2020). What is plurilingualism and what does it mean for language assessment? In G. Seed (Ed.), *Cambridge Assessment English Research Notes*, Issue 78 (pp. 5–15). Cambridge Assessment English.
- Shipley, K., & McAfee, J.G. (2020). Assessment in speech pathology (6th ed.). Plural Publishing.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331–345.
- Shohamy, E. (2001). The power or tests: A critical perspective on the uses of language tests. Pearson Education.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95(3), 418–429.
- Shohamy, E. (2013). A critical perspective on the use of English as a medium of instruction at universities. In A. Doiz, D. Lasagabaster, & J. Sierra (Eds.), *English-medium instruction at universities* (pp. 196–210). Multilingual Matters.
- Shohamy, E., & Pennycook, A. (2019). Extending fairness and justice in language tests. In C. Roeser & G. Wigglesworth (Eds.), *Social perspectives on language testing* (pp. 29–42). Peter Lang.
- Tatto, M. (2015). The role of research in the policy and practice of quality teacher education: An international review. *Oxford Review of Education*, 41(2), 171–201. <https://doi.org/10.1080/03054985.2015.1017405>
- Teacher Education Ministerial Advisory Group [TEMAG]. (2014). *Action now: Classroom ready teachers*. Retrieved from https://docs.education.gov.au/system/files/doc/other/action_now_classroom_ready_teachers_accessible.pdf
- Thompson, G., & Cook, I. (2014). Manipulating the data: Teaching and NAPLAN in the control society. *Discourse: Studies in the Cultural Politics of Education*, 35(1), 129–142.
- Thomson, S., De Bortoli, L., Underwood, C., & Schmid, M. (2019). *PISA 2018 in brief: Student performance*. Australian Council for Educational Research (ACER). <https://research.acer.edu.au/ozpisa/34>
- Thrupp, M. (2009). Teachers, social contexts and the politics of blame. *Queensland Teachers Professional Magazine*, 24, 6–12.
- Tupas, R. (Ed.). (2015). *Unequal Englishes: The politics of Englishes today*. Palgrave Macmillan.
- United Kingdom Department for Education. (2021). *Important: Changes to the professional skills test*. Retrieved from <http://sta.education.gov.uk/>
- van Gelderen, B. (2017). Growing our own: A 'two way' place-based approach to indigenous initial teacher education in remote Northern Territory. *Australian and International Journal of Rural Education*, 27(1), 14–28.
- Victorian Institute of Teaching [VIT]. (2021). *Literacy and Numeracy Test for Initial teacher education (LANTITE)*. Retrieved from <https://www.vit.vic.edu.au/faqs/lantite>
- Xinhua. (2020). China to cancel teachers' qualification exam for competent graduates. Xinhua. http://www.xinhuanet.com/english/2020-08/17/c_139297569.htm
- Zyngier, D. (2016). Education: *Unis treat teaching courses as a cash cow*. The Age. Retrieved from <http://www.theage.com.au/comment/the-age-letters/education-unis-treat-teaching-courses-as-a-cash-cow-20160122-gmc0py.html>