

The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities.

Susy Macqueen
Australian National University, Australia

Ute Knoch
University of Melbourne, Australia

Gillian Wigglesworth
University of Melbourne, Australia

Rachel Nordlinger
University of Melbourne, Australia

Ruth Singer
University of Melbourne, Australia

Tim McNamara
University of Melbourne, Australia

Rhianna Brickle
University of Melbourne, Australia

Abstract

All educational testing is intended to have consequences, which are assumed to be beneficial, but tests may also have unintended, negative consequences (Messick, 1989). The issue is particularly important in the case of large-scale standardised tests, such as Australia's National Assessment Program–Literacy and Numeracy (NAPLAN), the intended benefits of which are increased accountability and improved educational outcomes. The NAPLAN purpose is comparable to that of other state and national 'core skills' testing programs which evaluate cross-sections of populations in order to compare results between population sub-groupings. Such comparisons underpin 'accountability' in the era of population-level testing. This study investigates the impact of NAPLAN testing on one population grouping that is prominent in the NAPLAN results comparisons and public reporting: children in remote Indigenous communities. A series of interviews with principals and teachers documents informants' first-hand experiences of the use and effects of NAPLAN in schools. In the views of most participants, the language and content of the test instruments, the nature of the test engagement and the test washback have negative impacts on students and staff, with little benefit in terms of the usefulness of the test data. The primary issue is the fact that meaningful participation in the tests depends critically on proficiency in Standard Australian English (SAE) as a first language. This study contributes to the broader discussion of how reform-targeted standardised testing for national populations affects sub-groups who are not treated equitably by the test instrument or reporting for accountability purposes. It highlights a conflict between consequential validity and the notion of accountability which drives reform-targeted testing.

Key words

Test impact, literacy testing, numeracy testing, NAPLAN, accountability, global education reform movement, remote Indigenous communities, commensuration

Introduction

Educational reform movements arise in 'standard language cultures', where a language variety has taken hold economically and is imposed through institutional means (Milroy, 2001, 2012). Arguably, there is no more powerful means of imposing and maintaining a standard variety than standardized literacy testing, a phenomenon which has become increasingly widespread to the extent that comparisons using such tests within and between jurisdictions amount to 'global panopticism' (Lingard, Martino, & Rezai-Rashti, 2013, p. 539). At the same time, studies have documented detrimental effects of standardised testing on school curricula and language minority students (see, for example, David, 2011; Dulfer, Polese, & Turnbull, 2012; Mathison, 2003; Menken, 2008). Such tests are particularly problematic for assessing the academic achievement of children who are not proficient in the standard variety (Genishi & Brainard, 1995; Gutierrez-Clellen, 1996). The present paper is a case study in the issues represented by these global developments and practices. It deals with the impact of national standardized testing in Australia on the education of Indigenous children in remote communities where the first language of the children is generally not English and where English is first encountered on entry to school.¹

National literacy and numeracy testing in Australia

National standardized literacy and numeracy testing in Australia, known as the *National Assessment Program–Literacy and Numeracy* (NAPLAN), was introduced in 2008. The testing program replaced various State-based standardized tests and was introduced ahead of the new national curriculum. NAPLAN is designed to systematically measure and monitor the English literacy and numeracy of all children in years 3, 5, 7 and 9.² The test development and review process, accessible at <https://nap.edu.au/about/test-development>, includes details of the test and item development and review process, including the trialling which is conducted with representative student samples across all states and territories. The tests are also reviewed by an expert panel, before a second national sample of students takes part in an additional 'equating trial' which is carried out prior to the main test to ensure that the results can be reported on the same assessment scale.

Delivered annually over a three-day period, the tests assess four 'domains': numeracy, writing, reading, and language conventions.³ The numeracy test covers numerical and algebraic calculations, measurement and geometry, and statistics and probability. The writing test requires one composition which is either persuasive or narrative writing. The reading test includes different text types and samples skills ranging from finding directly stated information to identifying the tone of an argument. The language conventions test targets spelling, grammar and punctuation. Besides the written composition, most items across the tests are multiple choice with some constructed response items, e.g. write the correct spelling. Detailed descriptions of the tests, and sample papers, can be found at <https://www.nap.edu.au/naplan/the-tests>.

¹ The term 'Indigenous' refers to Aboriginal and Torres Strait Islander peoples, the original inhabitants of the Australian continent. It is the term used in official NAPLAN-related documents.

² Corresponding ages are approximately 8, 10, 12 and 14 years old.

³ <https://www.nap.edu.au/naplan> (Accessed 8 December, 2016)

The rationale for the development of NAPLAN arose from a general series of reforms in the 2000s which increasingly nationalised curriculum and assessment (for political context see Lingard & Sellar, 2013). The primary impetus for the development of a national standardized testing program was a declaration by the Ministers of Education in 2008 which inscribed national testing and an emphasis on literacy and numeracy (MCEETYA, 2008) as key means to achieving the broad goals of 1) 'Australian schooling promotes equity and excellence' and 2) 'all young Australians become successful learners, confident and creative individuals, and active and informed citizens' (n.p.). Prominent in the declaration was the intention to improve outcomes for Indigenous youth (to 'close the gap') and the provision of targeted support for this purpose. It was also explicit about the gathering of 'good quality data' about the performance of schools to enable comparison, and public reporting for 'strengthening transparency and accountability' (n.p.). According to the *Australian Curriculum and Testing Authority* (ACARA), the body responsible for NAPLAN development, administration and reporting, the tests are designed to 'determine whether or not young Australians are meeting important educational outcomes' (ACARA website, 2016). ACARA emphasises 'two benefits' of the NAPLAN tests: 1) to 'help drive improvements in student outcomes' and 2) to 'provide increased accountability for the community' (ACARA website, 2016).⁴

In relation to the first benefit, 'driving improvements', ACARA states:

All Australian schools benefit from the outcomes of national testing. Schools can gain detailed information about how they are performing, and they can identify strengths and weaknesses which may warrant further attention.

In relation to the second benefit, 'accountability', ACARA states:

Australians can expect education resources to be allocated in ways that ensure that all students achieve worthwhile learning during their time at school. The reported outcomes of the NAP [National Assessment Program] enable the Australian public to develop a general national perspective on student achievement and, more specifically, an understanding of how their schools are performing.

'Why NAP?' ACARA website⁵

These aims—*accountability* and *improvement*—are hallmarks of the Global Education Reform Movement (GERM), which typically uses standardized test instruments to implement educational policy (Chalhoub-Deville, 2016; Sahlberg, 2015). Prominent in a reform agenda, the concept of 'accountability' is unidirectional: teachers and schools are held accountable to governments and the populace through standardized test results that are made public through some means. As stated in the NAPLAN statement of benefit above, the test results provide the public with 'a general national perspective on student achievement' with a particular focus on 'how their schools are performing'. Therefore, the test is a powerful policy enactment which steers the 'systems and the work of schools' (Lingard et al., 2013, p. 539).

⁴ <https://www.nap.edu.au/about/why-nap> (Accessed 8 December, 2016)

⁵ <https://www.nap.edu.au/about/why-nap> (Accessed 2 November, 2016)

GERM testing samples cross-sections of populations for the purposes of comparisons between population sub-groupings, such as geographical areas, e.g. comparison of Australian states by NAPLAN in Lingard and Sellar (2013) or pre-identified social categories, e.g. comparisons of gender achievement by *Programme for International Student Assessment (PISA)* and *Pan Canadian Assessment Program (PCAP)* by Martino and Rezai-Rashti (2013). These types of comparisons are enabled through a process of 'commensuration' whereby different entities (e.g. schools, districts, states) and categorizations (e.g. geographical areas, gender, ethnicity) are compared by transferring their qualities into a common metric – the test scores (Espeland & Stevens, 1998; Lewis & Hardy, 2017). Commensuration underpins the notion of 'accountability' in the era of population-level educational testing (Hoffer, 2000); comparisons by category on test measures are how governments 'do' accountability. The explicit categorizations which are used to organise NAPLAN scores for accountability-based comparisons are state, sex, Indigenous status, Language Background Other than English status (LBOTE), parental occupation, parental education, and geolocation (metropolitan, provincial, remote and very remote).⁶ This type of data can be considered 'catalyst data' because it brings about the need for explanations for differences in test results across categories (Lingard & Sellar, 2013, p. 635). Thus, the accountability is ostensibly at the level of government, as it is with other testing programs that provide system-level comparisons, such as PISA. However, the publishing of NAPLAN school results on a website known as *MySchool* ensures that the results are highly visible at a local level.⁷ One use of the website is to enable 'fair and reasonable comparisons among schools with similar students' through a complex measure known as the 'Index of Community Socio-Educational Advantage' (ICSEA) which is calculated using geographical location, the proportion of Indigenous Students, and parents' education and occupation.⁸ Therefore the commensuration is carried out at the level of individual schools via the ICSEA measure, NAPLAN scores and the public website as a portal for accountability.

For language testers, a more familiar perspective on accountability has its roots in validity theory, which has, for nearly 50 years, included the consequences of testing as part of the discussion of test validation (Bachman, 2005; Chalhoub-Deville, 2009; Cronbach, 1988; Kane, 2013; Lynch, 2001; McNamara, 2008; Messick, 1989; Shohamy, 2001). The impact of high-stakes tests, defined as '...any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole' (Wall, 1997, p. 291), is the focus of accountability concerns for educational testers. In this view, the positive and negative effects of the test are part of the validation of the meaningfulness and usefulness of test scores (e.g. Fulcher & Davidson, 2007; Messick, 1996). This educational measurement view of accountability relates to the defensibility of the interpretation and use of the test instrument in relation to its effects on populations of test-takers. Those held accountable for negative consequences are test providers/developers and policy makers, i.e. those who *effect* the test instrument and policy (also see discussion of responsibility in Chalhoub-Deville, 2009). In contrast, the GERM notion of accountability is concerned with the defensibility of school and teaching quality in

⁶ In 2016 the decision was made to use the ABS Australian Statistical Geography Standard (ASGS) for NAPLAN (major cities, inner regional, outer regional, remote and very remote) <http://www.nap.edu.au/information/glossary> (Accessed 11 February, 2017)

⁷ <https://www.myschool.edu.au>

⁸ <https://www.myschool.edu.au/MoreInformation/About> (Accessed 9 January, 2017)

http://www.acara.edu.au/resources/About_icsea_2014.pdf (Accessed 9 January, 2017)

relation to the compared performances of particular sub-groups of test-takers. In reform-targeted testing, accountability flows downward so that those held accountable are schools, teachers and even test-takers and their parents, i.e. those who *are affected by* the test-policy (see also discussion in Lingard, Sellar, & Lewis, 2017).

In jurisdictions that implement educational reform using standardized tests, the process of commensuration provides a metric which combines persuasively with the metaphor of 'distance' (Lewis & Hardy, 2017), a readily-invoked policy hook, e.g. in the US, 'no child left behind' and 'race to the top', and, in Australia, 'close the gap', which refers to the difference in average scores of Indigenous Australian children and non-Indigenous children.⁹ An Australian policy application of this process is the current target to halve the English literacy and numeracy gap between Indigenous and non-Indigenous students by 2018 using 2008 NAPLAN results as baseline data (Council of Australian Governments, 2009). As Guenther observes, NAPLAN has become embedded in the 'rhetoric of schooling'; the notion of the 'gap' offers a definition of educational disadvantage, which, in turn, risks manufacturing the understanding that 'being Aboriginal or Torres Strait Islander is the disadvantage' (2013, pp. 157-8).

NAPLAN and children in remote Indigenous communities

The language ecologies of Australian remote Indigenous communities

The Indigenous population of Australia constitutes 3% of the total population. Of this Indigenous population, 22% live in remote or very remote communities scattered across (the north of) Western Australia, the Northern Territory, and Queensland (Australian Productivity Commission, 2014). In these remote communities, the language ecology is varied and complex with high rates of language shift underway (Marmion, Obata, & Troy, 2014), and the array of languages being spoken in any remote community is often complex. Children may be learning as their L1 one of several options. Some children will grow up learning the traditional language of the area in which they live, although only around 15-18 traditional languages continue to be learned by children as a first language. Other children may be living in a community where language is undergoing rapid shift and they are active in the process of developing new mixed languages (e.g. Gurindji Kriol Meakins, 2012; and Light Warlpiri O'Shannessy, 2005). Alternatively, children may be acquiring one of the many varieties of Kriol as their first language such as Roper River Kriol, Fitzroy Valley Kriol, Barkly Kriol or Wumpurrani English, another creole language such as Yumplatok or Yarrie Lingo, or one of a number of different Aboriginal English varieties (see discussions in Meakins, 2014; Sellwood & Angelo, 2013). What all these children have in common is that they are growing up in language environments where Standard Australian English (SAE) is not spoken and they are not acquiring SAE as their first language. Rather, their environment is much more akin to an English as a Foreign Language (EFL) environment than to that of migrant children who live in English speaking communities. In these remote communities, SAE is not generally spoken among members of the Indigenous community and is, for the most part, limited to some official government activities, health and justice contexts, the classroom and the local shop. For most children, their first,

⁹ In the case of literacy and numeracy, the 'gap' is determined by test scores, but the metaphor is used to describe a range of policy areas, (e.g. health, employment) each of which has an associated process or processes of commensuration.

more sustained exposure to SAE and to written forms of language comes when they start school at 5-6 years old.

NAPLAN and the monitoring of Indigenous children

Indigenous children are a sub-grouping of the Australian population who are identified in the test data and subject to explicit monitoring via the NAPLAN tests. This monitoring has consistently found that Indigenous children are not achieving at the same level as the rest of the population across all states (Ford, 2013). Angelo (2013) points out that the disaggregation of Indigenous students' scores has made this group 'highly visible' and hence subject to 'intensive educational interventions' on the basis of poor NAPLAN results, but with no recognition in the test data that many Indigenous children are also English language learners (p. 54, see also Dixon & Angelo, 2014).

The test results from the first administration in 2008 indicated that the test scores of Indigenous children were lower than non-Indigenous children, and that all children living in remote areas achieved lower scores than their urban peers (ACARA, 2008). Indigenous children in remote areas, however, scored lower than all remote students across all subtests at all grade levels (ACARA, 2008). Media reports at the time focused on the performance of Indigenous children, with headlines in major newspapers such as *One in Three Indigenous Kids Fail Test* (Ferrari, 2008) and *Indigenous Students Falling Further Behind* (Adlam, 2008). The impact of this pressure was greatest in the Northern Territory (NT), which has the highest population of Indigenous children living in remote areas, very few of whom are native speakers of Standard Australian English (SAE). The NT Government mandated that all state schools teach in English for the first four hours of each day—a policy which effectively saw the end of the few remaining bilingual education programs in place, even if 'programs had been running fairly successfully with strong community support' (Devlin, 2011, p. 263; Devlin, Disbray, & Devlin, 2017; Simpson, Caffery, & McConvell, 2009).

Since its introduction, there have been two government inquiries into NAPLAN (Education and Employment References Committee, 2014; Education, Employment and Workplace Relations References Committee, 2010), each of which attracted submissions concerning NAPLAN and Indigenous students. The 2010 report recommended that 'ACARA analyse and report publicly on how NAPLAN tests are serving different groups of Language Background Other Than English (LBOTE) students' as a response to the non-recognition of the language learner status of many Indigenous students (p. 18). However, in dealing specifically with Indigenous student performance, no mention was made of language learner status despite the reported committee view that 'ACARA should investigate ways to provide access to rich information on lower achievers' (p. 20), an action which would almost certainly reveal that English language proficiency is a prominent variable for many Indigenous children (see Angelo, 2013). The 2014 inquiry noted that NAPLAN contains material which presumes both English-as-a-first-language developmental levels and experience of urban environments, but the recommendations focussed predominantly on the introduction of a computer adaptive version of the test which was hoped to solve the gamut of complex issues raised.

Impact of NAPLAN

In the case of NAPLAN, the most widely reported effect or washback has been a narrowing of the curriculum through a greater focus on test task types and test practice (Dulfer et al., 2012; Hardy, 2015; Thompson, 2014). A narrowed curriculum echoes the experience of other jurisdictions with national testing programs (e.g. Berliner, 2011; Reay & Wiliam, 1999), and the narrowing can result in impoverished instruction for students for whom the tests are in a second language (Creagh, 2016; Menken, 2006, 2008). Narrowing of the curriculum and test task practice which exceeds the recommended test format familiarization have also been reported in schools with majority Indigenous children in Australia (Angelo, 2012, 2013; Unsworth, 2013). Unsworth's (2013, p. iii) broad survey of teachers' views of NAPLAN in 30 remote schools also identified a misalignment between the 'teaching to the test and teaching to Indigenous students' funds of knowledge'. Related to this misalignment is the identification of cultural inappropriateness of NAPLAN materials for children in remote communities (Unsworth, 2013; Wigglesworth, Simpson, & Loakes, 2011), for some of whom such foreign content could only compound the effect of the foreign language of the test.

The interaction between the writing test task and the second language learner status of some Indigenous children has also been the subject of research. An analysis of practice narratives for the NAPLAN writing test by children who are L1 speakers of Torres Strait Creole showed that the children's second language developmental level impacts their test responses, and further, that even the ability to 'fill up' overtly-taught narrative structures (e.g. introduction, problem, solution and conclusion) requires a level of L2 proficiency which is beyond the SAE competence of some of these children (Angelo, 2012, p. 50).

Method

Research aim

In response to the concerns documented in the literature and the high visibility of Indigenous children in the annual NAPLAN reporting and media response, this study investigates the impact of NAPLAN in remote Indigenous communities. Specifically, we posed two broad questions:

1. From the perspectives of school staff with experience of NAPLAN in remote Indigenous communities, what are the impacts of NAPLAN in remote Indigenous communities?
2. Are the reported effects positive or negative?

Participants

The interviewees were a purposive sample of 23 school staff who had combined experience of NAPLAN in 20 Indigenous communities classified as 'very remote' in Western Australia, Queensland and the Northern Territory.¹⁰ The category of 'very remote' in Australia is determined by distance to the nearest cities and towns. A very remote community may be hundreds of kilometres from the nearest urban centre.

¹⁰ <http://www.nap.edu.au/Information/Glossary#l> (Accessed 3 November, 2016)

The participants' teaching experience ranged from four to thirty years; 18 with experience as teachers and 5 with experience as teachers and principals. The small sample size reflects the proportion of teachers with experience of NAPLAN in remote communities which is also very small (as an indication, around 3% of all primary school teachers worked in remote or very remote areas in 2001).¹¹ Our selective sampling procedure matched our research question by seeking the perspectives of principals and teachers, both Indigenous and non-Indigenous, with first-hand experience of the use and effects of NAPLAN in remote Indigenous communities.

Procedure

The interviews were semi-structured and each lasted approximately 30 minutes. Interviews were carried out by members of the research team by phone, by Skype or face-to-face. Prior to the interviews, potential participants were informed that the interview

will focus on what happens at your school around the time of NAPLAN testing and whether or not NAPLAN has an effect on aspects of schooling such as curriculum, student attendance, student motivation, use of home languages and English, teaching and assessment practices and any other aspects you think are relevant. We will also be interested to hear how NAPLAN affects the community more generally and any changes you believe should be made to the tests.

The interview questions were broadly designed to explore any effects the test had on schooling (see questions in Appendix A). Participants were also invited to raise any topics they felt were relevant to the research focus. Our use of interviews is as means of collecting information, rather than a social practice, as distinguished in Talmy (2011). One clear disadvantage of this approach is that it treats interview data as individually constructed by the interviewee, rather than the collaborative effort brought about by the interview context and the semi-structured (non-standardized) nature of the questions (see, for example, discussions in Mann, 2011; Talmy, 2010). It should also be noted that NAPLAN is a charged topic among educationalists in Australia, with many teachers perceiving it to have negative effects on various aspects of schooling (Dulfer et al., 2012). That said, we believe it is in the interest of the broader community to hear the views, as far as we can present them, of those who have experienced a national standardized testing program in these complex linguistic environments.

The audio recordings were transcribed and analysed thematically using NVivo software by two coders in an iterative analytic process based on the processes set out in Miles, Huberman and Saldaña (2014). The first stage involved data annotations on the interview transcripts, as well as identification of main themes inherent in the interview questions. These annotations and expected themes formed the basis of broader themes or 'pattern codes' (Miles et al., 2014, p. 73). Pattern codes included practical considerations (e.g. *administration*), practices (e.g. *test strategies*), beliefs or perceptions (e.g. *cultural appropriateness*), and more theoretically-oriented topics (e.g. *usefulness of data*). These themes were trialed collaboratively by two analysts on a data subset, followed by coding of the whole dataset by one coder. The next stage involved a synthesis of these coded excerpts into 'higher level analytic meanings' by

¹¹ <http://www.abs.gov.au/ausstats/abs@.NSF/2f762f95845417aeca25706c00834efa/459c3882fad473a2ca2570eb0083be84!OpenDocument> (Accessed 26 October, 2017)

the second analyst (Miles et al., 2014, p. 95). In this stage, the pattern codes were related to theoretical aspects of test impact, (for example in Cheng, 2008; Messick, 1996; Wall, 1997), which formed the broad categories under which we present the findings below. These themes are inevitably selective due to the richness of the interview insights, the constraints of space and the need to consider the data in relation to the broader issues which are of relevance to other contexts in which national testing is carried out.

Findings

The findings are presented in three broad areas relating to test impact: i) Construct-irrelevant variance, ii) Psychosocial impacts iii) Use of NAPLAN results.

Construct irrelevant variance

Presumed English language proficiency. All participants agreed that the English language proficiency status (i.e. English as a second/foreign language or dialect) of Indigenous students in remote communities is a major barrier to meaningful participation in NAPLAN. As Teacher 4 observed, many children first encounter English at school when:

they've never seen books, they've never seen the language written down, they haven't seen their own language written down; they come from an oral background, they come in not even speaking English, and now we're getting them to sit down and try and do a NAPLAN test. Can you see how absurd that is?

Principal 1 described the enormity of the test English *literacy* demands in relation to students' *oral* English proficiency: 'When you struggle to speak a sentence in Standard Australian English, it's really hard to write a narrative.'

The writing test was generally perceived as posing the greatest challenge because meaningful participation requires a high level of written English language proficiency. Less meaningful participation was achieved in some cases through the use of strategic genre formulas for the NAPLAN text types (persuasive or narrative). Further, there was some disparity reported between the conditions and level of the writing test task and the typical process-based nature of classroom writing. For example, Teacher 9 noted that the writing task topic can be 'quite difficult to access' without 'support and negotiating the field of discussion before they get going'. Participants described various writing test strategies which compensate for students' inability to provide an engaged response. Teacher 3, for example, described students simply writing all the words they knew. Teacher 8 explained that she would tell the students to write a sentence they had learnt previously which 'might not have anything to do with the topic but they might have a full stop and a capital letter and that's something that they can put on the page'.

The numeracy test drew comment in relation to the word-based nature of many of the items, which was considered a barrier to demonstration of skills: the numeracy test is 'almost a test of English comprehension, rather than mathematical ability' (Teacher 10). One accommodation is that administrators may read aloud worded parts of the numeracy test (ACARA, 2015). However, it was reported that in practice, the accommodation made the test delivery time-consuming and incoherent

because the numerical component of the question cannot be read aloud: 'You're trying to read it to a child who's got English as a second language but leaving bits out' (Teacher 8).

Presumed cultural knowledge. Teachers reported instances in which the texts and items in both the numeracy and literacy tests presume experience and knowledge to be common across childhoods, but are 'totally alien' to the life experience of children in remote communities (Principal 2). Participants provided many examples of concepts which may be familiar to children in urban areas, but are relatively foreign in a remote community, for example the topic of 'recycling' (Principal 3), 'hosing things into gutters' (Teacher 10), tenpin bowling (Teacher 3) and going to the cinema (Teacher 14). Issues of cultural inappropriateness are perhaps most problematic in the writing test where the whole task hinges on a single concept. For example, 'changing a law', which as Teacher 15 explained, is a complex and unfamiliar topic in the context of an unfamiliar type of activity (the test).

Bubble-filling. Many participants referred to 'bubble-filling' – the multiple-choice response format widely used across the reading, language conventions and numeracy sections – as a test strategy or even an activity in and of itself, rather than a format for testing literacy or numeracy. Bubble filling gave the students something they could do 'even if they didn't understand the question, they were able to put something' (Principal 1). Teacher 4 reported telling the students 'to pick their favourite letter and fill it in'. Teacher 2 described the flurry of bubble-filling that occurred even 'before you say, turn your paper over'. Principal 2, in the excerpt below, indicates that for students with very limited English proficiency, there was little engagement with the test beyond bubble-filling:

It's just, some of them would say, 'what do I do here?'. And you would just say, 'well, fill in the bubble'. [...] if they were capable of reading you would say, 'read the question and fill in a bubble'. [...] if they weren't capable of reading I would just say, 'just fill in a bubble'.

Not surprisingly, there was a general perception that the test results were, to some extent, a product of guesswork.

Washback on classroom and curriculum

The interview data revealed a great deal of variability in the school practices regarding test preparation. Principal 2 explained that her school did 'nothing specific' in the lead up to the test except for familiarising the students with some of the question types. Teacher 3 described a clear belief in *not* doing specific test preparation tasks because, in her view, test practice was of 'no use' as such tasks do not engage the students' abilities. Similarly, Principal 1 felt the SAE speaker competence that underpinned the test made it unfeasible to prepare for the tests in any meaningful way: 'The tests themselves weren't going to cater for ESL [English as a Second Language] students [...] so you know, we were teaching the curriculum, we weren't trying to teach the test'. In contrast, Teacher 1 reported considerable amounts of test preparation, with teachers 'very focused on teaching to the test', particularly the NAPLAN text types.

Principal 3 described a lot of work on persuasive texts and weekly practice tests which he felt had lifted their results.

Teacher 11 observed two approaches to NAPLAN preparation. One was to do practice tests and the other was to create NAPLAN-format resources that link the test to the teaching program 'so they're getting an understanding of what a test is about with the content that they're familiar with'. An example of this more learner-based approach was to adapt a popular ESL resource so that it included NAPLAN-style multiple choice questions (Teachers 8, 9 and 10). Not surprisingly, multiple choice format was widely mentioned as a pre-taught feature of the test. Teachers 8 and 14 described 'practising colouring in multiple choice' so that the students didn't sit there doing nothing. Ironically, Teacher 16 felt that NAPLAN actually 'damaged their thirst for literacy and reading ... because we'd have to stop our regular literacy program that was working so well, for a week leading up and then do these tests and then go back to it and we've missed two weeks of quality teaching time.' This was echoed by Teacher 1 who felt that 'the amount of time that is spent just focusing on content that you think is going to be the test prevents kids accessing all areas of learning and getting those rich learning experiences.'

Psychosocial impacts of test experience

Psychological impact on students. There were various perceptions of how students were affected psychologically by the test. These ranged from having little impact to having strong negative effects, which may result from 'the emphasis that the school puts on the test and the individual teacher' (Teacher 1). Some teachers felt the test was simply irrelevant and an inconvenience. Teacher 17, for example, explained that for her Year 3s, the experience wasn't 'overwhelmingly positive or overwhelmingly negative, it was kind of an experience that for them was probably a bit nonsensical; they couldn't really understand what we're asking them to do or why we're asking them to do it.'

Some teachers observed strong negative effect on the students, in terms of stress or self-esteem. Teacher 14 explained that despite schools trying various methods of making the students feel supported and promoting the chances of success on the test, she could not see any aspect of the testing that suited the remote Indigenous context in which she taught. For her, the English language proficiency required for test participation raised issues of student wellbeing:

I didn't feel they needed that pressure on them because then they would just shut down completely. I had one child during NAPLAN hiding under the table, throwing pencils out, just stressed completely about it and because they didn't understand it, despite all the preparation. [...] I just let them in the end, write whatever they wanted for the writing section and they felt like they were succeeding and I let them believe they were succeeding and told them how proud I was of them and things like that because to me, the number one thing I'm looking after in the classroom is the wellbeing of the students [...] it was just being in English it was too hard for them.

Attendance is reportedly affected by NAPLAN, with students staying away on NAPLAN days, or experiencing one day of testing and then refusing to return for the other test days and reports that

parents kept their children away from school on NAPLAN days out of concern for the negative impact of the test pressure (Teachers 12, 14). While attendance is an issue throughout the school year in many remote communities, principals reported that they mobilised attendance officers for NAPLAN (Principal 1) or even collected children personally (Principal 2). Teacher 3, a NAPLAN administrator, explained that the sporadic attendance across the NAPLAN week meant that 'it would often take 3 or 4 times to actually finish it'. Staff also went to some effort to let the parents know attendance was especially important on NAPLAN days, for example going to the store or around the community to alert parents.

Psychological impact on teachers. Some teachers felt themselves to be negatively affected by the NAPLAN administration, which was 'possibly harder for the teachers than it is for the students because, as teachers, we know what the test is and we know what the implications for the results of the test are' (Teacher 17). Some teachers felt they themselves were being judged by the test results, while top-down pressures associated with NAPLAN results (described by Teacher 1 below) could affect the culture of the school:

I think NAPLAN affects the school and the students by actually changing the culture of the school. I think it's because there's been what's sold as a high-expectation agenda or an improvement agenda [...] and the only thing that is used to measure whether outcomes are improving for students is NAPLAN. So principals get pressure from their regions who get pressure from the central office who get pressure from wherever that comes from, from above, and then that impacts on principals pass that pressure down to teachers so that's almost changed the culture of teaching.

Teacher 5 explained that although 'it was kind of good to see what would be expected of kids in mainstream education', it was also 'disheartening' to see that her own students were 'so far behind', even though she acknowledged that English may be a third or fourth language for her students. For some, this resulted in a very negative emotional response, described by Principal 1 below:

...almost every year after the tests have finished [...] the teachers are generally feeling soul-destroyed, all feeling absolutely devastated [...] they feel so bad about putting the children through it.

Impact on student-teacher relationships. The test experience was perceived by several teachers as having a negative effect on the student-teacher relationship. Teachers felt that their inability to help students during the test eroded the trust that characterises student-teacher relationships. They felt responsible for imposing an impossible task on the students at the same time as refusing assistance which would normally be provided. For example, Teacher 9 described it as 'breaking that bond in that you're giving the students something that you know they can't do and you're not giving them any support'.

To counteract the negative effects on the students, there was also an element of protection expressed where teachers tried to insulate and compensate for the perceived risk to students' esteem. For example, teachers gave 'pep-talks' (Teacher 10) or 'played it down' (Teacher 11) as well as giving rewards or praise at the end of the test (Teacher 4). Teacher 14 explained that she was 'never honest with

the kids', preferring to make 'it seem like a fun thing' with them because she didn't want it to affect their self-esteem.

Use of NAPLAN results

Although participants indicated that they took note of NAPLAN results, it was widely expressed that the NAPLAN data has little relevance or usefulness for schools or parents. Comparisons between schools using the *MySchool* website were found to be unhelpful due to the small, variable cohorts described by Principal 1 above. Principal 4 said that the school used some of the results, but 'at the end of the day it doesn't show us a hell of a lot'. In addition, the very small, variable cohorts made the interpretation of data problematic, as suggested in the following excerpt:

we spent a lot of time looking at the data, and I still have nightmares sometimes about box and whisker data analysis with graphs [...] The Year 5s this year in 2007 did this and in 2009, look at this improvement, you can work on that! And you're trying to explain to them, look, they were completely different kids; we had a white kid in the class in 2009 and that's why the results [are] so much better. This is out of a class of about four Year 7's or something (Principal 1)

There were instances in which results were used in curriculum planning: in order to address weaknesses (Principal 3) or decide which areas of maths to cover (Teacher 10). Teacher 1 described sharing and analysing the results, looking at 'which year levels did well in what test [...] and what teachers they think had what impact on what results'. She perceived this process put 'a lot of pressure' on teachers.

At the level of the jurisdiction, Teacher 9 described how 'every year the NT government gets kicked' at the release of NAPLAN results. He explained that despite government reviews identifying bilingual education as the best evidenced-based practice, NAPLAN results and government-set NAPLAN goals put destructive pressure on bilingual programs.

Discussion

These data, from teachers and principals who have experienced firsthand the effects of NAPLAN in remote Indigenous communities, show that there is cause for concern about the appropriateness of test use in these contexts and the resulting effects of the mismatch between student and teacher needs on the one hand, and test instruments and usefulness of results on the other. On the whole, staff did not think that the test results reflected the abilities of the students for four main, interrelated reasons, the first of which is primary: 1) English language is not the first language and in some cases, not the first literacy of the children 2) the test content is inaccessible or foreign 3) a large proportion of children draw on construct-irrelevant test strategies such as bubble-filling 4) the tests do not engage the children cognitively or socioculturally. In the view of most participants, these characteristics of the test instruments (language/content) and test-taker behaviours (test engagement) have negative psychosocial impacts on students and staff, with little or no benefit in terms of the usefulness of the test data. The reported washback on classroom and curriculum amounts to superficial test-survival strategies which

potentially even damage students' literacy development, rather than promote it (see also Angelo, 2013). In the views of the participants in this study, the NAPLAN experience and effects appear to be, at best, an irrelevance and at worst, the source of considerable distress. This finding echoes other surveys in the Australian context that have concluded that most teachers feel the effects of NAPLAN have been largely negative (Dulfer et al., 2012; Thompson, 2014).

These effects are compounded by the very high visibility of Indigenous children in the national analysis, which disaggregates and discusses this group's performance annually. However, arguably, not including Indigenous children in remote communities in a national testing program would separate this ethnocultural group in ways that are also potentially detrimental. It is important to note that participants were not opposed to standardized testing *per se*, but rather to the presumption of English native-speaker language competence and western urban culture that they perceived to underpin test constructs. Connected to this is the lack of discernment at the lower end of attainment; as Teacher 9 explained, his students were 'pretty much all deemed to fail equally' even though he knew they were not all at the same level.

A fundamental issue is that the NAPLAN test constructs are underpinned by English-as-a-first-language literacy (Angelo, 2013; Creagh, 2016). Writing from the US context, Solano-Flores (2009) observes that the inclusion of English language learners in national testing 'has not been accompanied by effective approaches to minimize construct-irrelevant score variance due to language issues' (p. 33). He characterises the testing of such learners as a 'stochastic process' due to uncertainty about variables such as sensitivity to contextual features, and language of the test delivery (Solano-Flores, 2009, p. 43). As observed by the participants in this study, even the numeracy test is rendered inaccessible through its heavily English-language construct. To exemplify this, of the 120 items on the four Year 3 numeracy tests from 2008 to 2011, there were just two items composed only of numerical symbols or shapes. All other items require an understanding of written English for meaningful engagement and, in many instances, these are complicated syntactic structures, such as 'In which game did Kim score 3 more goals than she scored in Game 2?' (NAPLAN Year 3 Numeracy Test, 2008). The NAPLAN constructs, therefore, are a poor fit with the language ecologies of many remote Indigenous communities, where there is limited presence of print-based literacy in any language. Indigenous children in remote communities have generally not acquired any of the pre-literacy skills in either English or their home language/s that their urban, English-speaking counterparts are typically equipped with in readiness for schooling (Kral, 2012). This has a profound effect on their acquisition of English academic literacy, and inevitably, on their performance on NAPLAN.

Messick (1980) reminds us that the question of test fairness arises when an individual's group membership determines that he or she achieves significantly differently from the population as a whole. As measured by NAPLAN, the significantly different performance of remote Indigenous students has been sustained and highly visible in reporting and media in every test administration. The quantification provided by NAPLAN remains prominent in public discourse about 'closing the gap', with annual revisits in the media at NAPLAN reporting time which pore over the unclosed gap (e.g. 'Reforms Failing to Close Indigenous Schools Gap', Ferrari, 2012). In addition to the effects that these participants reported in relation to students, teachers, classroom and curriculum, the commensuration, i.e. the comparison

different groupings on a common metric (Espeland & Stevens, 1998), provided by NAPLAN has become absorbed in Australian educational and public discourse; it has become 'common sense' (Star & Lampland, 2009). An easily-comprehended metric, NAPLAN belies the nature of schooling in complex sociolinguistic contexts, thus rendering an 'empirically intricate reality deceptively straightforward' for the purpose of comparison (Stevens, 2008, p. 102). Thanks to this systemic absorption, 'the numbers themselves come to be regarded as coextensive with the underlying phenomena they measure' (Stevens, 2008, p. 102). This contribution of test-user perspectives has interrogated this assumption and highlighted the complexity underpinning the metric.

Conclusion

Standardized testing programs such as NAPLAN are intended to bring about change in the literacy level (as an indicator and proxy for 'educational outcomes') of the population. The test enacts a policy underpinned by efforts to increase life opportunities at an individual level (e.g. Mahuteau, Karmel, Mavromaras, & Zhu, 2015) and a correlation between literacy and economic growth at a national level (e.g. Coulombe, Tremblay, & Marchand, 2004). The body responsible for NAPLAN, ACARA, has stressed that the use of NAPLAN results acts as impetus in this causal chain below:

It should be emphasised that NAPLAN is a tool to inform school improvement, not an improver of educational outcomes. It is not the tests that will improve students' literacy and numeracy skills, but the way students' results (including school, system and national level results) are used by teachers, schools and systems to identify strengths and weaknesses, particularly in teaching practices and programs, that will improve student outcomes.

(ACARA submission, reported in *NAPLAN Inquiry Report*, Education and Employment References Committee, 2014, p. 58)

As discussed earlier, the test results are 'catalyst' for the accountability process which applies, according to ACARA, 'particularly in teaching practices and programs'. Yet, this method of accountability is not necessarily a democratic one. Shohamy points out that, although standardized testing 'may be based on equal opportunities and conditions of testing', it does not necessarily follow that 'all test takers have equal opportunity to obtain an equal education' (2004, p. 87). Indeed, she lists the imposition of tests that are 'based on one type of knowledge' on test-takers from different backgrounds as a possible violation of test-taker rights. In relation to Indigenous children in remote communities, it is unrealistic, as Kral (2009, 2012) points out, to assume that the development of academic English literacy can be achieved in a foreign language through brief pedagogical interventions to a level comparable to that of middle-class Australia (see also Angelo, 2013). Arguably, practices we have come to recognise as indications of fairness in educational testing such as 'careful standardization of tests and administration conditions' in the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) may well be the cause of ill-effect when the context, the educational environment and the test-taker linguistic background differs so markedly from the literacy practices valued by the 'standard language culture' (Milroy, 2001)

and on which the construct is modelled. According to Angelo and Dixon (2014), this has resulted in a widespread ignorance of the value of multilingualism or need for English language learning for Indigenous children. In their view, Australia's 'monolingual mindset' (Clyne, 2008; see also Shohamy, 2011) has assisted the safe passage of NAPLAN 'literacy and numeracy testing' into schooling as if it were not in any language and hence fair, even for students who are speakers of Indigenous traditional and contact varieties, with little English proficiency. As Guenther (2013) concludes, 'the assumptions behind the administration of testing instruments, the language used in them and the values embedded within the tests should be questioned' (p. 167). If an identifiable sub-population of test-takers are unable to engage with test content because of their proficiency in the language of delivery, the test has poor 'interactiveness', in Bachman and Palmer's (1996) test usefulness framework. Not being able to engage with the test instrument precludes meaningful results and thus also the kind of reform envisaged by policy makers. The 'interactiveness' observed by our interviewees in the NAPLAN appeared to be limited to filling in the bubbles with no sense of whether or not the response was correct.

This brings us back to the proposition that accountability in educational reform runs counter to accountability in consequential validity. Shohamy's view, based on the notion of consequential validity in educational measurement, is that 'there should be accountability with regard to the purpose of the test, its practice and its methods' (2004, p. 87). Those responsible for the design and use of the test – developers, providers, policy makers – need to ensure that the instrument does no harm (Kunnan, 2000, 2004). It is clear, however, that the teachers and principals in this study feel the onus is on them to defend the poor performances of their students in a GERM sense of accountability. Caught between a powerful policy instrument and the needs of their students, many participants also felt the need to protect their students from the unintended negative effects of the test instruments. These findings suggest that this process of commensuration—the quantitative comparison of the qualities of different sub-groups—is not informative about or beneficial for children in remote Indigenous communities. For comparison to be carried out, then there is a clear need for an instrument which is sensitive to the development of literacy and numeracy in English as a foreign language which is accessible for these students. At a local level, such an instrument should be able to discern learning in these contexts in a way that might be informative for teaching and beneficial for learning. At a national level, an instrument that recognises the multilingual capacity of many Australian children has the potential to disrupt the myth of monolingualism in Australia's classrooms.

References

- Adlam, N. (2008). Indigenous students falling further behind. *Northern Territory News/Sunday Territorian*.
- Angelo, D. (2012). *Sad stories: A preliminary study of NAPLAN practice texts analysing students' second language linguistic resources and the effects*. Paper presented at the Proceedings of the 42nd Australian Linguistic Society Conference - 2011.
- Angelo, D. (2013). NAPLAN implementation: Implications for classroom learning and teaching, with recommendations for improvement. *TESOL in Context*, 23(1/2), 53.
- Australian Curriculum and Reporting Authority. (2015). *NAPLAN 2015 National protocols for test administration*: Downloaded from: <http://nap.org.au/naplan/school-support/national-protocols.html> (Accessed 17 November, 2015).
- Australian Curriculum Assessment and Reporting Authority. (2008). *2008 National Assessment Program Literacy and Numeracy Achievement in Reading, Writing, Language Conventions and Numeracy*. Retrieved from Sydney: <https://www.nap.edu.au/results-and-reports/national-reports> https://www.nap.edu.au/resources/2ndStageNationalReport_18Dec_v2.pdf
- Australian Productivity Commission. (2014). *Overcoming Indigenous disadvantage: Key indicators 2014*. Canberra: Australian Productivity Commission.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302.
- Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics*, 29, 118-131.
- Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language testing*, 33(4), 453-472.
- Cheng, L. (2008). Washback, impact and consequences. In N. H. Hornberger & E. Shohamy (Eds.), *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment* (pp. 349-364). Berlin: Springer.
- Clyne, M. (2008). The monolingual mindset as an impediment to the development of plurilingual potential in Australia. *Sociolinguistic Studies*, 2(3), 347-365.
- Coulombe, S., et al. (2004). *International Adult Literacy Survey: Literacy scores, human capital and growth across fourteen OECD countries*. Ottawa: Statistics Canada.
- Council of Australian Governments. (2009). *National Indigenous Reform Agreement*. Retrieved from Canberra: http://www.federalfinancialrelations.gov.au/content/national_agreements.aspx
- Creagh, S. (2016). A critical analysis of the Language Background Other Than English (LBOTE) category in the Australian national testing system: a Foucauldian perspective. *Journal of Education Policy*, 31(3), 275-289.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- David, J. (2011). What students need to learn: high-stakes testing narrows the curriculum. *Educational Leadership*, 68(6), 78-80.
- Devlin, B. (2011). The status and future of bilingual education for remote Indigenous students in the Northern Territory. *Australian Review of Applied Linguistics*, 34(3).
- Devlin, B., et al. (2017). *History of Bilingual Education in the Northern Territory* (Vol. 12). Singapore: Springer Singapore.
- Dixon, S., & Angelo, D. (2014). Dodgy data, language invisibility and the implications for social inclusion. *Australian Review of Applied Linguistics*, 37(3), 213-233.
- Dulfer, N., et al. (2012). *The experience of education: The impacts of high stakes testing on school students and their families, An educator's perspective*. Retrieved from Sydney:
- Education and Employment References Committee. (2014). *Effectiveness of the National Assessment Program – Literacy and Numeracy*. Retrieved from Canberra: http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/Naplan13/Report/index
- Education, Employment and Workplace Relations References Committee. (2010). *Administration and Reporting of NAPLAN Testing*. Retrieved from Canberra:

- http://www.aph.gov.au/sitecore/content/Home/Parliamentary_Business/Committees/Senate/Education_Employment_and_Workplace_Relations/Completed_inquiries/2010-13/naplan/report/index
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual review of sociology*, 24(1), 313-343.
- Ferrari, J. (2008, December 23). One in three Indigenous kids fail test. *The Australian*.
- Ferrari, J. (2012, December 19). Reforms failing to close Indigenous schools gap. *The Australian*. Retrieved from <http://www.theaustralian.com.au/national-affairs/education/reforms-failing-to-close-indigenous-schools-gap/news-story/c7b6c260d2d9eee60a49982824382470>
- Ford, M. (2013). Achievement gaps in Australia: What NAPLAN reveals about education inequality in Australia. *Race Ethnicity and Education*, 16(1), 80-102.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon: Routledge.
- Genishi, C., & Brainard, M. (1995). Assessment of bilingual children: A dilemma seeking solutions. In E. E. Garcia & B. McLaughlin (Eds.), *Meeting the Challenge of Linguistic and Cultural Diversity in Early Childhood Education* (pp. 49-63). New York: Teachers College Press.
- Guenther, J. (2013). Are we making education count in remote Australian communities or just counting education? *The Australian Journal of Indigenous Education*, 42(02), 157-170.
- Gutierrez-Clellen, V. (1996). Language diversity: Implications for assessment. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of Communication and Language* (pp. 29-56). Baltimore: Paul H. Brookes.
- Hardy, I. (2015). A logic of enumeration: The nature and effects of national literacy and numeracy testing in Australia. *Journal of Education Policy*, 30(3), 335-362.
- Hoffer, T. B. (2000). Accountability in education. In M. T. Hallinan (Ed.), *Handbook of the Sociology of Education* (pp. 529-543): Springer.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association. <http://www.apa.org/science/programs/testing/fair-testing.pdf> (Accessed 17 January, 2017).
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kral, I. (2009). *The literacy question in remote Indigenous Australia*. Canberra: Centre for Aboriginal Economic Policy Research, Australian National University.
- Kral, I. (2012). *Talk, text and technology: Literacy and social practice in a remote Indigenous community*. Bristol: Multilingual Matters.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (Vol. 9, pp. 1-14). Cambridge: University of Cambridge, Local Exams Syndicate.
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.), *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference, July 2001* (Vol. 18, pp. 27-48). Cambridge: Cambridge University Press.
- Lewis, S., & Hardy, I. (2017). Tracking the topological: The effects of standardised data upon teachers' practice. *British Journal of Educational Studies*, 65(2), 219-238.
- Lingard, B., et al. (2013). Testing regimes, accountabilities and education policy: commensurate global and national developments. *Journal of Education Policy*, 28(5), 539-556.
doi:10.1080/02680939.2013.820042
- Lingard, B., & Sellar, S. (2013). 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28(5), 634-656.
- Lingard, B., et al. (2017). Accountabilities in schools and school systems. In G. Noblit (Ed.), *Oxford Research Encyclopedia of Education*. Oxford: Oxford University Press.
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language testing*, 18(4), 351-372.
- Mahuteau, S., et al. (2015). *Educational Outcomes of Young Indigenous Australians*. Adelaide: Report submitted to the National Centre for Student Equity in Higher Education (NCSEHE), Australia. National Institute of Labour Studies (NILS), Flinders University.
- Mann, S. (2011). A critical review of qualitative interviews in applied linguistics. *Applied Linguistics*, 32(1), 6-24.
- Marmion, D., et al. (2014). *Community, Identity, Wellbeing: The Report of the Second National Indigenous Languages Survey*. Retrieved from Canberra: <http://aiatsis.gov.au/publications/products/community-identity-wellbeing-report-second-national-indigenous-languages-survey>

- Martino, W., & Rezai-Rashti, G. (2013). 'Gap talk' and the global rescaling of educational accountability in Canada. *Journal of Education Policy*, 28(5), 589-611.
- Mathison, S. (2003). The accumulation of disadvantage: The role of educational testing in the school career of minority children. *Workplace*, 10, 32-43.
- McNamara, T. (2008). The socio-political and power dimensions of tests. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7: Language testing and assessment, pp. 415-427). Dordrecht: Springer.
- Meakins, F. (2012). Which mix—code-switching or a mixed language?—Gurindji Kriol. *Journal of Pidgin and Creole Languages*, 27(1), 105-140.
- Meakins, F. (2014). Language contact varieties. In H. Koch & R. Nordlinger (Eds.), *The Languages and Linguistics of Australia: A comprehensive guide* (pp. 365-416). Berlin: De Gruyter Mouton.
- Menken, K. (2006). Teaching to the test: How *No Child Left Behind* impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521-546.
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon: Multilingual Matters.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3 ed., pp. 13-103). New York: American Council on Education and Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language testing*, 13(3), 241-256.
- Miles, M. B., et al. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks: Sage.
- Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, 5(4), 530-555.
- Milroy, J. (2012). Sociolinguistics and ideologies in language history. *The handbook of historical sociolinguistics*, 569-584.
- Ministerial Council on Education, Employment Training and Youth Affairs. (2008). *Melbourne declaration on educational goals for young Australians*. Retrieved from Melbourne:
- O'Shannessy, C. (2005). Light Warlpiri: A new language. *Australian Journal of Linguistics*, 25(1), 31-57.
- Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343-354.
- Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Sellwood, J., & Angelo, D. (2013). Everywhere and nowhere: Invisibility of Aboriginal and Torres Strait Islander contact languages in education and Indigenous language contexts. *Australian Review of Applied Linguistics*, 36(3), 250-266.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Longman.
- Shohamy, E. (2004). Assessment in multicultural societies: Applying democratic principles and practices to language testing. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 72-92). Cambridge: Cambridge University Press.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95(3), 418-429.
- Simpson, J. H., et al. (2009). *Gaps in Australia's Indigenous language policy: Dismantling bilingual education in the Northern Territory* (0855756713). Retrieved from
- Solano-Flores, G. (2009). The testing of English language learners as a stochastic process: Population misspecification, measurement error, and overgeneralization. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research* (pp. 33-45). New York: Routledge.
- Star, S. L., & Lampland, M. (2009). Reckoning with standards. In M. Lampland & S. L. Star (Eds.), *Standards and their stories: How quantifying, classifying and formalizing practices shape everyday life* (pp. 3-24). Ithaca: Cornell University Press.
- Stevens, M. L. (2008). Culture and education. *The ANNALS of the American Academy of Political and Social Science*, 619(1), 97-113.
- Talmy, S. (2010). Qualitative interviews in applied linguistics: From research instrument to social practice. *Annual Review of Applied Linguistics*, 30, 128-148.
- Talmy, S. (2011). The interview as collaborative achievement: Interaction, identity, and ideology in a speech event. *Applied Linguistics*, 32(1), 25-42.
- Thompson, G. (2014). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing. *International education journal: comparative perspectives*, 12(2).
- Unsworth, P. (2013). *Studying the effects of NAPLAN on Indigenous education in remote schools, Australia*. (PhD Thesis), School of Education, University of South Australia.

- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 291-302). Dordrecht: Kluwer Academic Publishers.
- Wigglesworth, G., et al. (2011). NAPLAN language assessments for Indigenous children in remote communities: Issues and problems. *Australian Review of Applied Linguistics*, 34(3), 320-343.

Appendix A Interview questions

A. INTERVIEWS FOR PRINCIPALS

Thank you for taking the time to participate in this study. As you know we are looking at NAPLAN and what effects it has on schools – the students, the teachers and the community more generally.

Could you give me a general picture of how NAPLAN affects your school and the students?

Could you describe the NAPLAN testing period here?

Does NAPLAN testing have any effect on teaching practices? (If so, what...)

What is your opinion of the NAPLAN test papers and tasks (in relation to the students here)?

Would you recommend any changes to NAPLAN?

Do you feel pressure about your school's performance on NAPLAN?

How do you think your students and teachers view NAPLAN?

B. FOCUS GROUPS FOR TEACHERS

Could you give me a general picture of how NAPLAN affects your school and the students?

Could you describe the NAPLAN testing period here?

How do you think your students feel about NAPLAN?

Do you feel pressure about your students' performances on NAPLAN?

Does NAPLAN testing have any effect on what you do in class? (If so, what...)

What is your opinion of the NAPLAN test papers and tasks (in relation to the students here)?

Would you recommend any changes to NAPLAN?



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Macqueen, S; Knoch, U; Wigglesworth, G; Nordlinger, R; Singer, R; McNamara, T; Brickle, R

Title:

The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities

Date:

2019

Citation:

Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T. & Brickle, R. (2019). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing*, 36 (2), pp.265-287. <https://doi.org/10.1177/0265532218775758>.

Persistent Link:

<http://hdl.handle.net/11343/217031>

File Description:

Accepted version