NAPLAN discourses: a systematic review after the first decade

Judy Rose ^{a,b}, Samantha Low-Choy^a, Parlo Singh^b and Daniela Vasco^b

^aOffice of Research, Faculty of Arts, Education and Law Group, Griffith University, Brisbane, Australia; ^bGriffith Institute for Educational Research, School of Education and Professional Studies, Griffith University, Brisbane, Australia

ABSTRACT

In 2018 NAPLAN (National Assessment Program – Literacy and Numeracy) reached a 10-year milestone. Introduced in 2008 by the Australian Curriculum, Assessment and Reporting Authority (ACARA), NAPLAN assesses student literacy and numeracy in Years 3, 5, 7 and 9. As a key education reform in Australia, NAPLAN has provoked critical debate across public, political and academic fora. This study synthesises themes emerging from a quantitative systematic review of NAPLAN-related literature to identify the main discourses that have emerged in response to this large-scale standardised assessment programme. The dominant discourses coalesce around ‘datafication’, ‘social justice’, ‘affect and emotion’, and ‘accountability and performativity’. The stakes are high in these discursive struggles around NAPLAN as large-scale standardised testing is reconstituting the purposes of education, the professionalism and professional identities of educators, and affecting parents and students’ engagement in schooling.

KEYWORDS

Large-scale standardised testing; NAPLAN; systematic quantitative review; data discourses; datafication of education; big data

Introduction

It is a decade since NAPLAN (National Assessment Program – Literacy and Numeracy) was introduced into all Australian schools as a significant education reform. This large-scale standardised test, administered by the Australian Curriculum, Assessment and Reporting Authority (ACARA), is undertaken annually by students in Years 3, 5, 7 and 9. NAPLAN is mandatory across all school systems, and federal funding is contingent on participation. NAPLAN standardised test results were designed as an overall indicator of literacy and numeracy performance by testing skills deemed important for the twenty-first century (ACARA, 2011).

NAPLAN signals an important education reform in Australia, through the adoption of education policy orientated to standardisation of assessment with a distinctly Anglo-American agenda (Gable & Lingard, 2016). Indeed, NAPLAN can be conceived as an Australian manifestation of an international trend to institute national testing. As noted by Sahlberg (2006), the Global Education Reform Movement (GERM), has included centralised proscription of assessment as well as the attendant curriculum; central monitoring of assessment; and competitive reporting about school performance. Lingard, Thompson, and Sellar (2016) argue the ‘focus of GERM is on

top-down, test-based modes of accountability linked to parental choice and market reforms, including the restructuring of the bureaucracy, and a whole variety of privatisations' (p. 2).

In Australia, through the generation of data that seeks to compare, order and rank schools, NAPLAN has helped to facilitate and constitute a move to a national system of schooling (Lingard et al., 2016). The 'utopian' promise is that measurement of learning through standardised testing at different scalar levels (national and international) 'will help to improve efficiency and quality (for example system, school, teacher and student performance) and that more and better data will give insights into the often-opaque world of the classroom' (p. 1). In turn, this stance is driving change globally so that the 'purposes of education are now narrowly conceived as the production of a certain quantity and quality of human capital ... leading to the "economisation" of education policy' (p. 1). While some commentators have suggested that national testing is a conservative force, narrowing curriculum and scripting teaching, Thompson, Sellar, & Lingard, 2016 (cited in Lingard et al., 2016) argue that it can be seen as a progressive force in two distinct ways:

First, the introduction of testing is set against a narrative of technical and administrative progress: that is, testing enables more rational and efficient production of knowledge about, and administration of, education systems. Second, the primary justification for the introduction of NAPLAN and MySchool – the suggestion that it would enable more equitable and redistributive allocation of resources and help to improve educational outcomes – has appeal to groups that can be seen as progressive in their political orientations. (p. 217)

As the debates around NAPLAN have morphed over the last ten years, it is timely to discern how NAPLAN 'serves as a communicative device between the world of education and wider society' (Broadfoot & Black, 2004, p. 9). This is especially pertinent at a point in time when NAPLAN transitions into an adaptive computer-based, online environment (Thompson, 2017). The term 'adaptive' or 'tailored' testing signals that 'online tests will be responsive in real time providing an integrated personalised testing, pedagogy and intervention for each student' (Thompson et al., 2016, p. 212).

Here we review existing literature to reveal how NAPLAN is framed by 'set[s] of practices, statements and "truths" (discourses) that privilege certain interpretations, values and expectations above others' (Thompson & Lasic, 2011, p. 5). Starting with a comprehensive systematic review to discover relevant literature, we seek to identify themes that emerge via a discourse analysis on NAPLAN.

Rationale for NAPLAN

Introduced by the Rudd/Gillard Labor government in 2008, the NAPLAN school reform initiative received bi-partisan political support from inception. Thus, as a signatory to the Melbourne Declaration on the Educational Goals for Young Australia, the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) committed to promote an international standard of curriculum and testing by collecting 'good quality' data that would be 'world class' to meet accountability goals (MCEETYA, 2008). This shift in 'policy focus on international comparative organisational performance' aimed to establish Australia's world ranking (Gable & Lingard, 2016). It emulates the US *No Child*

Left Behind policy: a tri-pronged policy of academic standards, standardised testing and economic accountability (Smith, 2014). The National Assessment Program (NAP) includes NAPLAN and PISA (Programme for International Student Assessment) with each operating as components of a wider data network (Lingard et al., 2016).

Implementing NAPLAN has required state-level education jurisdictions to relinquish control over assessment to public-sector management, with such testing overseen by ‘edu-businesses’ (Thompson, Savage, & Lingard, 2015, p. 2), thus producing a ‘quasi-market’ in education (Leckie & Goldstein, 2017, p. 194). These hybrid edu-business entities function within such educational markets by doing new forms of educational work, e.g. test preparation, remedial services and products, data management and analysis. NAPLAN, and other large-scale tests like PISA, have been designed to fit this market model of social and economic accountability.

Policy rhetoric proclaimed NAPLAN would allow parents to make informed choices about children’s education. As then Minister for Education, Gillard (2010) announced – the public display of NAPLAN results would herald ‘A new era of transparency – *MySchool* has put information in the hands of parents and created a new national conversation about what we should expect from our schools and about what it takes to deliver great results’ (p. 3). It was thought that parents would gain deeper insights into their child/ren’s progress through personalised reports and be able to compare their child’s school with similar schools or against the national average via the *MySchool* website (Mills, 2015). Yet, the graphical and numerical summaries supplied on *MySchool*, are complex and prone to misinterpretation and misuse (Ladwig, 2010; Wu, 2010). For instance, *MySchool* NAPLAN reports are often misinterpreted as an indication of school quality (Smith, 2014) and misused to produce league tables (Thompson & Lasic, 2011).

What does NAPLAN measure?

The main design aim of NAPLAN was to provide standardised and hence comparable measures of student achievement in English literacy (with four main tests: reading, writing, spelling, grammar /punctuation) and in numeracy (with one test covering algebra, geometry, measurement and problem solving) (ACARA, 2011). Standardisation allows the students in a particular year to be compared with other years (Marks, 2015, p. 128).

To date, there has been no comprehensive review of NAPLAN-related literature, although reports on NAPLAN by several ‘think tank’ organisations have included substantial literature sections. However, this consultative work has been critiqued in terms of issuing simplistic solutions to complex issues in schooling, including standardised testing (Loughland & Thompson, 2016); as noted, ‘This explains the tendency for think-tank organisations to ignore context, complexity and nuance in favour of recipes’ (p. 115).

Other scholars concur that empirical evidence on the effects of NAPLAN is lacking or limited (Ragusa & Bousfield, 2017) particularly detail on the degree of distress on teachers, students and parents (Rogers, Barblett, & Robinson, 2016): ‘NAPLAN/ *MySchool* [has] garnered negative media coverage and stakeholder criticism, although little research empirically examined their impact of children’s lived experiences’ (pp. 266–267). This motivates our review of a more extensive range of empirical evidence on NAPLAN.

Methodology of review

We conducted a Systematic Quantitative Review (SQR) to ‘identify as much of the relevant literature as possible, but not extend too far into less pertinent fields’ (Pickering & Byrne, 2014, p. 543). We chose SQR as it is more comprehensive than other literature review typologies, which may be unspecified, narrative, scoping or critical (Grant & Booth, 2009). The comprehensive search necessarily enumerates a ‘sampling frame’, permitting *quantification* of the prevalence of themes in the literature (Pickering & Byrne, 2014). Casting a wide search net using multiple databases (see Figure 1, step 1), we subsequently screened for depth, breadth and quality (see Figure 1, step 2), then appraised the literature for eligibility and inclusion (see Figure 1, steps 3–4) before synthesising evidence in depth in following sections.

Figure 1. PRISMA flow diagram – NAPLAN literature.

Note: Key Search Terms: ‘NAPLAN’, ‘assessment’, ‘testing’ AND ‘national testing’, ‘data’, ‘teaching’ AND ‘pedagogy’ AND ‘curriculum’, ‘MySchool’, ‘literacy’ Informit database search (44 databases): Including AEIPT (A + Education), ERIC (Education Resources Information Centre), Pro Quest, Web of Science, Scopus, Social Sciences Citation Index. Source: PRISMA, adapted from Moher, Liberati, Tetzlaff, and Altman (2009). Preferred Reporting for Systematic Reviews and Meta-Analyses: The PRISMA Statement.

In our analysis of the literature four main overlapping themes, listed in order of coverage were: ‘Datafication’ (46%), ‘Social Justice’ (37%), ‘Affect and Emotion’ (31%), and ‘Accountability and Performativity’ (26%) (See Table 1). Our selection of terminology (like datafication, accountability, performativity) reflects the ways in which NAPLAN has been discursively constructed, and organisation of literature around these thematic labels is our original contribution. These thematic headings were created by using a hybrid interpretivist/phenomenological approach (See Rose, 2017). Discursive themes emerging from the literature were classified as phenomena (e.g. datafication), or concepts with sub-constructs (e.g. Social Justice – disadvantage, diversity and equity). This analysis achieved a balance of themes (around 26%–35% each), and offset the potential limitations inherent in a more quantitative review approach. However, this study offers but one of the ways to construct a discursive account of NAPLAN.

Discourses of NAPLAN data

Inevitably, NAPLAN-related literature encompasses issues of assessment, pedagogy (teaching) and curriculum, i.e. the knowledge field in which NAPLAN is embedded. Thus, we do *not* identify these topics as separate discursive themes. Yet, there is common concern in the identified literature over a mis-alignment between NAPLAN test items and discipline-specific, curriculum knowledge transmitted to students (Frawley & McLean Davies, 2015). Similarly, we do *not* identify ‘the policy implications of NAPLAN’ as a separate theme, as they are prevalent, ubiquitous and inform the broader debates, as noted a decade after national educational testing arrived in the UK:

The importance of policy speaks for itself. Firstly, decisions about assessment procedures – particularly those concerning ‘high-stakes’ testing of various kinds – are as often based on *political* appeal as they are on a systematic knowledge of the scientific evidence concerning fitness for purpose. (Broadfoot & Black, 2004, p. 9)

The four key themes will now be discussed in more detail.

Table 1. Key themes across NAPLAN literature.

THEMES – Sub-themes	Exemplar authors*	% Coverage **
DATAFICATION		46%
Datafication	Marjanovic and Cecez-Kecmanovic (2017); Smith (2018)	
Big data	Chesters and Daly (2017); Pugh and Foster (2014)	
SOCIAL JUSTICE		37%
Disadvantage	Loughland and Thompson (2016)	
Diversity	Creagh (2014); Wigglesworth, Simpson, and Loakes (2011)	
Equity	Cumming and Dickson (2013); Dixon and Angelo (2014)	
AFFECT & EMOTION		31%
Affect	Sellar (2015); Rogers et al. (2016)	
Emotion	Howell (2017); Swain, Pendergast, and Cumming (2018)	
ACCOUNTABILITY & PERFORMATIVITY		26%
Accountability	Klenowski and Wyatt-Smith (2012); Smith (2014)	
Performativity	Lupton and Williamson (2017); Singh (2017)	

*Authors listed are examples, not representative of all literature identified.

**Percentages indicates the coverage or spread of the literature.

1. Datafication

A strong theme relates to discourses of 'datafication' (digital practices in education), and the impact on teaching and learning. Processes of datafication in education have worked in parallel with policy to achieve an auditable record of student learning outcomes using large-scale, standardised performance data (Ball, 2003).

1.1. Datafication of education

The term *datafication* has emerged in an era of 'big data' and the accompanying growth of national and international standardised testing (Marjanovic & Cecez-Kecmanovic, 2017). The practices of datafication in education involve digital technologies that oversee, monitor and regulate the work of educators through the use of comparative algorithms and thereby 'reframing professional identities and relationships' (Singh, 2017, p. 6). Yet, data in the form of numbers, statistics and ranking are not new to education but 'central to state functioning since the rise of the nation-state in the eighteenth century' (Lingard, 2011, p. 356).

However, datafication practices in education have escalated with the proliferation of data and burgeoning digital technologies. Now, 'digitised forms of data are exerting enormous leverage in education institutions and policy, creating languages, practices, subjectivities and intensifications of power and force through reconfiguring of intensities, or "speeds" of institutional life' (Thompson, 2017, p. 828). This datafication process suggests that NAPLAN data may have become more powerful than intended.

A complex phenomenon arises in discourses of metrics, measures, indicators and targets, which act to describe and produce 'calculable worlds' and new governance sites and systems (Gorur, 2015, p. 581). Discourses of data gain their power and authority by presenting quantitative data as objective, accurate and reliable indicators of learning and teaching (Gable & Lingard, 2016). Moreover, datafication is closely tied to value systems or ideologies that ascribe it with particular qualities (e.g. trust/mistrust, truth/fiction) (Smith, 2018). Yet, irrespective of its form, qualitative or quantitative, data provide only glimpses of the 'truth' (Lingard, 2011, p. 358).

1.2. Perverse and positive effects

The 'high stakes' outcomes of NAPLAN has led to perverse effects, including intensive test preparation and/or the surreptitious exclusion or coaching of students to improve school results (Thompson & Cook, 2014). In addition, there are concerns about teachers and schools being judged on published results (Klenowski & Wyatt-Smith, 2012). In contrast, on a positive note, making NAPLAN results open and accessible through the *MySchool* portal could empower the public through actions of citizen democracy (Marjanovic & Cecez-Kecmanovic, 2017).

For some schools NAPLAN results have become catalysts for positive changes (Mills, 2015). They have helped target collaborative partnerships to boost literacy teaching and learning in disadvantaged schools through professional development (e.g. Singh, Märtsin, & Glasswell, 2015). NAPLAN results combined with diagnostic testing (e.g. PAT-M) information has been used to improved mathematics pedagogy and student numeracy outcomes in other schools (Muir, Livy, Herbert, & Callingham, 2018). Thus, a balance of both positive and perverse effects can be noted in relation to NAPLAN testing.

1.3. Big data

Although NAPLAN results are technically not large enough to be classed as ‘big data’, they still suffer many of these general issues. Large datasets such as NAPLAN and PISA are sometimes recognised as ‘big data’ (Pugh & Foster, 2014, p. 261), although this interpretation depends on discipline. NAPLAN may alternatively be seen as ‘open performance data’, as governments make it available to public view and citizen interpretation (Marjanovic & Cecez-Kecmanovic, 2017). Furthermore, big data are associated with analytic capabilities that make sense of unstructured digital information including learning analytics (Casanovas, De Koker, Mendelson, & Watts, 2017).

The new on-line NAPLAN testing involves computers that can produce ‘machine learning’ algorithms to predict test questions to match an individual student’s ability (Thompson, 2017). However, depending on how it is implemented, Australia’s choice to individually tailor NAPLAN tests online could contribute to a conceptual gap compared to its overseas counterparts. This technologically-driven innovation raises the stakes of national testing to the next level, via a pipeline of educational data analytics. This large expansion in *variety* of information, will necessarily increase its *volume* and hence demands on underlying infrastructure, as spectacularly demonstrated by first online delivery and subsequent ‘Australian census debacle’ (Casanovas et al., 2017, p. 339).

In addition, growing concern has been expressed by school leaders in regard to perceived barriers preventing a fuller understanding of the inner workings and logics of NAPLAN (Ladwig, 2010; Ragusa & Bousfield, 2017). The lens of Complex Adaptive Systems (CAS) is one suggested way to look into the ‘black box’ of open performance data and to examine their potential datafication effects (Marjanovic & Cecez-Kecmanovic, 2017). However, such meanings and intentions of data can become distorted as CASs evolve over time, and often in unpredictable ways.

1.4. Interpreting data like NAPLAN

Correct and appropriate interpretation of NAPLAN data is beginning to gain greater attention. Yet, ‘intelligent public debate is hampered by lack of public reporting on the technical characteristics of NAPLAN and a lack of public education on how to understand such data’ (Ladwig, 2010, p. 39). Thus, making sense of the technical aspects of NAPLAN remains a point of tension, particularly in cases where stakeholders may have limited statistical understanding and therefore be unable to fully appreciate the caveats in the results (Wu, 2016, cited in Lingard et al., 2016). Furthermore, poorer performing schools report concerns about how they should understand and productively work with NAPLAN data (Cormack & Comber, 2013).

Several cautions signal the care required to interpret NAPLAN, e.g. within the bounds of reported statistical accuracy (e.g. Wu, 2010). Measurement error is large (12%) for individual students, so that NAPLAN scores may be unreliable for diagnosing specific student learning difficulties (Ladwig, 2010; Wu, 2010). Other critics concur that ‘NAPLAN is not a diagnostic test, given its large-scale application to large groups of students to extrapolate how schools, not students, are performing’ (Harris et al., 2013, p. 34). Yet, this contrasts with political promises of NAPLAN’s diagnostic value to parents and teachers (ACARA, 2011; Masters, 2010). When these concerns were probed during Senate inquiries into NAPLAN and *MySchool*, custodians claimed NAPLAN was never intended as a diagnostic

test (ACARA, 2011). Ongoing criticisms point out that NAPLAN's public and political functions overshadow its diagnostic capabilities (Ladwig, 2018). Moreover, this issue has been difficult to resolve while NAPLAN is centrally controlled and politically supported.

1.5. Issues of validity, reliability and robustness

Several validity and reliability concerns have developed for NAPLAN. Standardisation is achieved via Rasch scaling, which delivers a relative *not* absolute indication of student achievement, despite usual interpretations as the latter (Tan, 2013). Specifically, Rasch scaling is more useful for determining which test questions (of varying difficulty) differentiate students rather than evaluating student ability. Wu, 2016 (cited in Lingard et al., 2016) notes the large margin of error in individual student performance relates to the limited number of test questions that provide only an indication of student ability. Further estimates on the calibrated equivalence scales for each year level, and domain of literacy or numeracy, are imprecise (Ladwig, 2018). Yet, these validity and reliability issues have not been clearly communicated by ACARA.

In terms of comparisons, improvements within a school could be masked by cross-sectional analysis (in specific years) but could be revealed via longitudinal analyses that track student improvement across years (Tan, 2013). When students change schools, their apparent NAPLAN 'improvement' may relate to individual or school socio-economic status (SES) or other factors (Marks, 2015). Analysing individual NAPLAN results from one territory, Chesters and Daly (2017) found that 'peer effects' (where school has proportionally fewer students in the lowest SES) increased academic achievement of students in Years 7 and 9.

Further debates around NAPLAN robustness, reliability and external validity have donned various methodological perspectives. For example, internal flaws were identified in the spelling domain of NAPLAN literacy testing (Daffern, Mackenzie, & Hemmings, 2017). The cultural relevance of written composition components was criticised for presuming similar experiences (e.g. access to cinemas) in remote versus middle-class, metropolitan settings (Frawley & McLean Davies, 2015).

NAPLAN numeracy results are potentially unreliable measures of mathematical ability, as they rely on strong language skills and cognitive capacities such as 'working memory', and hence 'a student who fails a numeracy test because of insufficient literacy abilities – rather than lack of numeracy knowledge or skill – is unlikely to benefit from numeracy remediation' (Howard, Woodcock, Ehrich, & Bokosmaty, 2017, p. 109). Others question NAPLAN's capacity to assess higher order cognitive skills such as creativity and critical thinking, so crucial to child and adolescent brain development (Swain et al., 2018).

2. Social justice

Discussed in 37% of publications reviewed, social justice issues address political and policy promises surrounding NAPLAN: to help to address issues of inequality in schooling and more equitably distribute funding for increased support and resources to schools and students with highest need (ACARA, 2011; MCEETYA, 2008). Yet, achieving these equity goals has been a challenge thwarted by the underlying market-driven principles and logics of comparative competition that are intrinsic to a large-scale standardised test such as

NAPLAN, and the tendency to increase rather than lessen inequities among schools (Ragusa & Bousfield, 2017).

2.1. Disadvantage

Evidence shows that non-participation in NAPLAN is highly correlated with conditions of poverty (e.g. prior contact with Child Protection agencies, history of parental mental health problems, and Indigenous status) (Wong, O'Donnell, Bayliss, Fletcher, & Glauert, 2017). Thus, ensuring children living in disadvantaged school communities are visible to funding allocations and increased support has been highly challenging. As Cormack and Comber (2013) note, 'One key problem with the discourse of data and its various technologies, is that the materiality of poverty and place is backgrounded' rather than afforded the nuance required (p. 87).

The main mechanism NAPLAN uses for recognising disadvantage is the school Index of Community Socio-Educational Advantage (ICSEA) score, derived from measures of parental education and occupation, school geographic location and the proportion of Indigenous students. ICSEA scores are dichotomised to define higher and lower SES schools, with the rationale it will enable 'fair' and 'meaningful' comparisons among schools with similar students (ACARA, 2011, p. 1). However, primary school principals report that the ICSEA scale is not 'fine-tuned enough' to capture disadvantage or diversity (Klenowski & Wyatt-Smith, 2012, p. 72).

2.2. Diversity

Another mechanism NAPLAN uses for recognising diversity is: Indigenous or LBOTE (Language Background Other Than English) categorisation of students. However, NAPLAN fails to recognise diversity for students with multi-lingual capabilities in non-standard forms of Australian English including 'Kriol' (Wigglesworth et al., 2011, p. 321). Designed for Standard Australian English speakers, NAPLAN is 'linguistically and culturally unsuitable for Indigenous children, especially for those living in remote communities' (p. 340). There is evidence that NAPLAN and *MySchool* have dampened debates about Indigenous languages in schools, and effectively eliminated bi-lingual/bi-cultural education in the Northern Territory (Gannon, 2013). Vass & Chalmers, 2016 (cited in Lingard et al., 2016) argue that NAPLAN, with its narrow focus on testing literacy and numeracy, has undermined the enactment of policies which have a decolonising potential, such as the Embedding Aboriginal and Torres Strait Islander Perspectives in Schooling (EATSIPS) programme.

Another concern is students can become 'lost' within the broad LBOTE (Language Background Other Than English) category (Ford, 2013). LBOTE is unable to differentiate recently arrived refugee students and hence renders them invisible for targeted funding and support (Creagh, 2014). Moreover, attention to addressing the gap between Indigenous and non-Indigenous student performance on NAPLAN, which is recognised as 'scandalous', needs to entail recognition of institutional racism, and for governments to consider 'radically different ways of providing education' (Ford, 2013, p. 83). These strategies might include teaching skills and competencies to enable disadvantaged students to succeed in the mainstream while also responding to needs of the community (Mills, 2015; Singh et al., 2015).

2.3. Equity

Equity has attracted strong comment. In rural Australia, school choice via *MySchool* has no practical meaning or benefit (Gannon, 2013). Further, students with disabilities may be deprived of the right to NAPLAN participation, and adequate support for additional needs (Cumming & Dickson, 2013). Some scholars argue that NAPLAN has inadequate procedures for inclusion, and further discriminates against high-learning-needs students, by labelling them academic failures (Dixon & Angelo, 2014).

Another paradox is despite the original intent to identify high-risk students, there is limited and variable support for their participation (Cumming & Dickson, 2013). This lax practice seems to contradict stated equity intentions of inclusivity across students (Gannon, 2013). One hypothesis is that a 'performative culture' of standardised testing is less likely to invest in students with additional needs or 'where the margins from improved performance are limited' (Ball, 2003, p. 224).

3. Affect and emotion

'Affect' and 'emotion' are concepts found in 31% of papers. These themes relate to how participants and stakeholders feel (respond emotionally) and are affected (in terms of work practices) by the testing regime.

3.1. Bodily affect

Ball (2003) had predicted the 'shockwaves' of large-scale standardised testing on teachers' sense of professional identity. Similarly, Sellar (2015) observed that first reactions to NAPLAN results are often bodily and visceral and involve complex emotions. In schools identified as 'failing', anxiety levels among staff, students and parents can be extremely high and are often accompanied by fear of public 'naming and shaming' (Cormack & Comber, 2013, p. 84). A rural school reported that while teachers understood the importance of gathering data through large-scale testing, they felt the assumptions underlying NAPLAN did not transfer well to remote school settings (Cormack & Comber, 2013). Therefore, teachers found it burdensome as it constricted the breadth of their pedagogy and compelled them to devote more time teaching to the test. In more advantaged schools, the effects of NAPLAN on students and staff were less acute, although still perceived to place undue pressures on teachers (Rogers et al., 2016).

3.2. Student attitudes and emotions

Some studies have specifically examined the effect of NAPLAN on student wellbeing (Howell, 2017; Lingard et al., 2016; Swain et al., 2018). Findings suggest children in low SES public schools frequently reported negative emotions regarding NAPLAN, drawing faces that showed fear, anxiety and sadness (Swain et al., 2018). In contrast, independent Catholic schools found more variation in children's responses to NAPLAN with negative responses mainly related to its unclear purpose and expressed as frustration, anger and boredom (Howell, 2017). Howell also found Year 7 students had more negative attitudes towards NAPLAN than younger cohorts, concerned it would affect their progression through school and into higher education. However, in schools where principals and teachers deliberately downplayed the importance of NAPLAN, negative emotions were less prevalent among students and staff (Muir et al., 2018).

4. Accountability and performativity

Issues of ‘Accountability’ and ‘performativity’ of NAPLAN arose in 26% of publications. Scholars acknowledged that NAPLAN’s accountability purposes were not unique to Australia, but part of a global movement in education (Hursh, 2007; Leckie & Goldstein, 2017; Lingard & Sellar, 2013).

4.1. Accountability apparatus

Globally, accountability discourses are associated with an array of ‘performance monitoring machinery’, and an ‘audit culture’ that employs various ‘actors’ to oversee, inspect, and scrutinise school performance by standardised testing (Ozga, 2016). Tensions have arisen as the accountability standards set by NAPLAN conflict with teachers trying to maintain pedagogical agency and positive relationships with students and colleagues (Harris et al., 2013). Heightened concerns around ‘accountability’ have flared when poor test performance is used to unfairly target and blame schools (e.g. Smith, 2014). Moreover, with use of accountability mechanisms such as *MySchool* there is risk that the public reporting of school NAPLAN results may penalise students who are the least advantaged (Ford, 2013; Mills, 2015).

The flawed logic of *MySchool* for teachers is that public display of NAPLAN data will make them work harder (Thompson & Cook, 2014). However, surveillance of students, teachers and schools through the construction of public data profiles, instead raises concerns about ‘dataveillance’, privacy and the potential violation of children’s learning rights (Lupton & Williamson, 2017). Therefore, students and teachers face different vulnerabilities and risks when NAPLAN data is reported publicly.

4.2. Performativity and identity

Performativity relates to the performances of teachers, students and schools in response to a raft of accountability measures (e.g. NAPLAN results) that are used to evaluate effectiveness, worthiness, integrity, quality and competence (Ball, 2003; Hursh, 2007; Leckie & Goldstein, 2017). Ball questions whether this pressure to perform means teachers are more likely to be ‘ontologically insecure’ and professionally apprehensive as they try to meet expectations that are often unclear (Ball, 2003, p. 220). Of added concern are ways in which a culture of performativity has shaped pedagogy – by reorganising teachers’ time and work practices and shifting the focus away from more desirable pedagogies (Ragusa & Bousfield, 2017). Moreover, despite an established level of professional autonomy, there is evidence to suggest teachers are increasingly measured and monitored against student performance on NAPLAN and PISA (Gorur, 2015; Singh, 2017).

Discussion

This paper has reviewed NAPLAN and its discursive impact, 10 years since implementation. Through this systematic review we have identified key literature, then synthesised the dominant discourses assembled in our analysis as: datafication, social justice, affect and emotion, and accountability and performativity. We sought to address an overarching question of how well NAPLAN serves as a communicative device for reporting students’ academic achievement in the key learning areas of literacy and numeracy. The evidence

shows that miscommunications, misunderstandings and misinterpretations are evident in regard to NAPLAN's purpose, function and value across different fora and among various stakeholder groups.

There are also recognised benefits of NAPLAN including the foregrounding of necessary life skills and targeted funding and resourcing to schools with the highest need (Singh et al., 2015). NAPLAN has helped to expose the 'scandalous' gap between Indigenous and non-Indigenous educational achievement outcomes, though it does little to address the complex issue of institutional racism (Ford, 2013). Further, it does not recognise the rich diversity in non-standard dialects of English, or encourage the teaching of Aboriginal languages (Wigglesworth et al., 2011). Conversely, evidence suggests that NAPLAN works best when used as one of several indicators of achievement, one tool in a toolbox to inform (not dictate) pedagogy and curriculum, and where the 'high stakes' aspect of the suite of tests is downplayed (Muir et al., 2018). As NAPLAN provides an annual snapshot of student learning, it can only provide 'vignettes of student achievement' rather than a detailed portfolio of learning progress over time (Cumming & Dickson, 2013, p. 10).

Brown and Hattie (2012) have further argued that standardised tests can be beneficial to education when: (1) they provide information on strengths as well as learning gaps, (2) teachers and students are well informed and have a strong belief in their purpose, and (3) they are conducted in an 'environment of psychological safety' (pp. 288–290). Our synthesis reveals that points 2 and 3 are not satisfactorily being met for NAPLAN. Moreover, New Zealand employs a different system of governance so that standardised test data are owned and controlled by schools rather than a central authority (Brown & Hattie, 2012). This issue of control may be crucial as NAPLAN continues to change as it enters a digital, on-line environment (Thompson, 2017).

Conclusion

NAPLAN's approach to national educational testing is at once subject to international trends whilst remaining highly specific to the Australian context. It essentially contributes a unique instantiation of national 'experiment' into the global arena of large-scale standardised assessment. In contrast to NAPLAN, UK national testing includes 'expected' student attainment into their standardisation, with some demonstrated bias towards particular types of schools (Ozga, 2016). This ostensibly produces school league tables, which have evolved to explicitly measure school performance and determine punitive actions, such as closure or takeovers, although transparency and fit-for-purpose are still being contested (Leckie & Goldstein, 2017). In similar vein, national testing in the US is strongly linked to meeting performance and accountability targets, with serious outcomes for teachers in 'failing schools' (Hursh, 2007, p. 297).

As NAPLAN is an evolving phenomenon, new questions and discourses regarding its meaning and adequacy of comparisons and ranking among students, and indeed schools, continue to emerge. The challenges of adaptive testing have brought forth fresh concerns in regard to the validity and reliability of data generated (Ladwig, 2018) and renewed consideration of assessment alternatives (Gonski et al., 2018). Of concern is that this move to online testing will be accompanied by intensified performativity and accountability pressures, given assurances of greater *velocity* and *veracity* (pertinence

to individuals and hence schools), fulfilling the four Vs hype of Big Data (Casanovas et al., 2017). Nevertheless, NAPLAN will likely invite further interdisciplinary dialogue as it enters another phase of review.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Australian Research Council Discovery Scheme [grant number DP160102784].

ORCID

Judy Rose <http://orcid.org/0000-0001-6383-7666>

References

- ACARA. (2011). *National protocols for test administration*. Melbourne, VIC: Australian Curriculum Assessment and Reporting Authority.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18, 215–228.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7–26.
- Brown, G. T., & Hattie, J. (2012). The benefits of regular standardized assessment in childhood education. In S. Suggate, & E. Reese (Eds.), *Contemporary debates in childhood education and development* (pp. 287–292). New York: Routledge.
- Casanovas, P., De Koker, L., Mendelson, D., & Watts, D. (2017). Regulation of big data: Perspectives on strategy, policy, law and privacy. *Health and Technology*, 7(4), 335–349.
- Chesters, J., & Daly, A. (2017). Do peer effects mediate the association between family socio-economic status and educational achievement? *Australian Journal of Social Issues*, 52(1), 63–77.
- Cormack, P., & Comber, B. (2013). High-stakes literacy tests and local effects in a rural school. *Australian Journal of Language and Literacy*, 36(2), 78–89.
- Creagh, S. (2014). A critical analysis of problems with the LBOTE category on the NAPLAN test. *The Australian Educational Researcher*, 41(1), 1–23.
- Cumming, J. J., & Dickson, E. (2013). Educational accountability tests, social and legal inclusion approaches to discrimination for students with disability: A national case study from Australia. *Assessment in Education: Principles, Policy & Practice*, 20(2), 221–239.
- Daffern, T., Mackenzie, N. M., & Hemmings, B. (2017). Testing spelling: How does a dictation method measure up to a proofreading and editing format? *Australian Journal of Language and Literacy*, 40(1), 28–45.
- Dixon, S., & Angelo, D. (2014). Dodgy data, language invisibility and the implications for social inclusion. *Australian Review of Applied Linguistics*, 37(3), 213–233.
- Ford, M. (2013). Achievement gaps in Australia: What NAPLAN reveals about education inequality in Australia. *Race Ethnicity and Education*, 16(1), 80–102.
- 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.
- Gable, A., & Lingard, B. (2016). NAPLAN data: A new policy assemblage and mode of governance in Australian schooling. *Policy Studies*, 37(6), 568–582.

- Gannon, S. (2013). My school redux: Re-storying schooling with the *MySchool* website. *Discourse: Studies in the Cultural Politics of Education*, 34(1), 17–30.
- Gillard, J. (2010, April 15). *A future fair for all: School funding in Australia: Address to the Sydney Institute*.
- Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, P. L., & Roberts, M. (2018). *Through growth to achievement: Report of the review to achieve educational excellence in Australian schools*. Canberra: Commonwealth of Australia.
- Gorur, R. (2015). Producing calculable worlds: Education at a glance. *Discourse: Studies in the Cultural Politics of Education*, 36(4), 578–595.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.
- 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.
- Howard, S. J., Woodcock, S., Ehrich, J., & Bokosmaty, S. (2017). What are standardized literacy and numeracy tests testing? Evidence of the domain-general contributions to students' standardized educational test performance. *British Journal of Educational Psychology*, 87(1), 108–122.
- Howell, A. (2017). 'Because then you could never ever get a job!': Children's constructions of NAPLAN as high-stakes. *Journal of Education Policy*, 32(5), 564–587.
- Hursh, D. (2007). Exacerbating inequality: The failed promise of the *No Child Left Behind* Act. *Race Ethnicity and Education*, 10(3), 295–308.
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65–79.
- Ladwig, J. G. (2010). What NAPLAN doesn't address (but could, and should). *Professional Voice*, 8(1), 35–40.
- Ladwig, J. G. (2018). Critique: On the limits to evidence-based learning of educational science. In *The Wiley handbook of teaching and learning* (pp. 639–659).
- Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'. *British Educational Research Journal*, 43(2), 193–212.
- Lingard, B. (2011). Policy as numbers: Ac/counting for educational research. *The Australian Educational Researcher*, 38(4), 355–382.
- 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., Thompson, G., & Sellar, S. (Eds.). (2016). *National testing in schools: An Australian assessment*. New York: Routledge.
- Loughland, T., & Thompson, G. (2016). The problem of simplification: Think-tanks, recipes, equity and 'turning around low-performing schools'. *The Australian Educational Researcher*, 43(1), 111–129.
- Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, 19(5), 780–794.
- Marjanovic, O., & Cecez-Kecmanovic, D. (2017). Exploring the tension between transparency and datification effects of open government IS through the lens of complex adaptive systems. *The Journal of Strategic Information Systems*, 26(3), 210–232.
- Marks, G. N. (2015). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education*, 41(1), 122–144.
- Masters, G. (2010, August). NAPLAN and *MySchool* shedding light on work in progress. *Teacher: The National Education Magazine*, 22–25.
- Mills, C. (2015). Implications of the *MySchool* website for disadvantaged communities: A Bourdieuan analysis. *Educational Philosophy and Theory*, 47(2), 146–158.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). (2008). *Melbourne declaration on educational goals for young Australians*. Melbourne: Ministerial Council on Education, Employment, Training and Youth Affairs.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.

- Muir, T., Livy, S., Herbert, S., & Callingham, R. (2018). School leaders' identification of school level and teacher practices that influence school improvement in national numeracy testing outcomes. *The Australian Educational Researcher*, 45(3), 297–313.
- Ozga, J. (2016). Trust in numbers? Digital education governance and the inspection process. *European Educational Research Journal*, 15(1), 69–81.
- Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development*, 33(3), 534–548.
- Pugh, K., & Foster, G. (2014). Australia's national school data and the 'big data' revolution in education economics. *Australian Economic Review*, 47(2), 258–268.
- Ragusa, A. T., & Bousfield, K. (2017). 'It's not the test, it's how it's used!' Critical analysis of public response to NAPLAN and MySchool Senate Inquiry. *British Journal of Sociology of Education*, 38(3), 265–286.
- Rogers, S. L., Barblett, L., & Robinson, K. (2016). Investigating the impact of NAPLAN on student, parent and teacher emotional distress in independent schools. *The Australian Educational Researcher*, 43(3), 327–343.
- Rose, J. (2017). "Never enough hours in the day": Employed mothers' perceptions of time pressure. *Australian Journal of Social Issues*, 52(2), 116–130.
- Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, 7(4), 259–287.
- Sellar, S. (2015). A feel for numbers: Affect, data and education policy. *Critical Studies in Education*, 56(1), 131–146.
- Singh, P. (2017). Performativity, affectivity and pedagogic identities. *European Educational Research Journal*, 17(4), 489–506.
- Singh, P., Märtsin, M., & Glasswell, K. (2015). Dilemmatic spaces: High-stakes testing and the possibilities of collaborative knowledge work to generate learning innovations. *Teachers and Teaching*, 21(4), 379–399.
- Smith, G. J. D. (2018). Data doxa: The affective consequences of data practices. *Big Data and Society*, 5(1), 1–15.
- Smith, W. C. (2014). The global transformation toward testing for accountability. *Education Policy Analysis Archives*, 22(116), 1–30.
- Swain, K., Pendergast, D., & Cumming, J. (2018). Student experiences of NAPLAN: Sharing insights from two school sites. *The Australian Educational Researcher*, 45(3), 315–342.
- Tan, J. (2013). *Mathematical modelling and statistical analysis of school-based student performance data* (Master's thesis). The University of Adelaide, Adelaide.
- Thompson, G. (2017). Computer adaptive testing, big data and algorithmic approaches to education. *British Journal of Sociology of Education*, 38(6), 827–840.
- 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.
- Thompson, G., & Lasic, T. (2011). *Doing something about it: Representations of NAPLAN in the public domain*. Retrieved from <http://eprints.qut.edu.au/86168/>
- Thompson, G., Savage, G., & Lingard, B. (2015). Think tanks, edu-business and education policy: Issues of evidence, expertise and influence. *The Australian Educational Researcher*, 43(1), 1–13.
- Thompson, G., Sellar, S., & Lingard, B. (2016). The life of data: Evolving national testing. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 212–229). New York: Routledge.
- Vass, G., & Chalmers, G. (2016). NAPLAN, achievement gaps and embedding Indigenous perspectives in schooling. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 139–151). New York: Routledge.
- Wigglesworth, G., Simpson, J., & Loakes, D. (2011). NAPLAN language assessments for Indigenous children in remote communities. *Australian Review of Applied Linguistics*, 34(3), 320–343.
- Wong, J. W., O'Donnell, M., Bayliss, D., Fletcher, J., & Glauert, R. (2017). Patterns of participation in Year 9 academic testing and factors predicting absence on the day of test. *Educational Research for Policy and Practice*, 16(2), 109–127.

- Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.
- Wu, M. (2016). What national testing data can tell us. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 36–47). New York: Routledge.

NAPLAN discourses: a systematic review after the first decade

Judy Rose, Samantha Low-Choy, Parlo Singh & Daniela Vasco

To cite this article: Judy Rose, Samantha Low-Choy, Parlo Singh & Daniela Vasco (2018): NAPLAN discourses: a systematic review after the first decade, *Discourse: Studies in the Cultural Politics of Education*, DOI: [10.1080/01596306.2018.1557111](https://doi.org/10.1080/01596306.2018.1557111)

To link to this article: <https://doi.org/10.1080/01596306.2018.1557111>

Published online: 07 Dec 2018.

Submit your article to this journal [↗](#)

Article views: 2

View Crossmark data [↗](#)