**The pros and cons of standardised testing, focusing on NAPLAN**

**Michael Murray (2019)**

Standardised tests are a feature of neoliberal educational reform policies in countries around the world. Governments and educational authorities, in particular, use the results of these tests to make judgements about the performance of schools in enacting policies at the local level. In discussing the reasonableness of using test data to judge the quality of individual schools, I will focus on the National Assessment Program – Literacy and Numeracy (NAPLAN), a standardised test that has been operating in Australia since 2008.

In an educational paper reviewing standardised testing in OECD countries, Morris (2011) defines standardised tests as ‘tests that are designed externally and aim to create conditions, questions, scoring procedures and interpretations that are consistent across schools’ (p5). Such tests are large-scale, undertaken simultaneously by great numbers of students, either full cohort (census-based) or significant samples. Morris identifies Australia’s national testing regime, NAPLAN, as a standardised test.

A significant feature of NAPLAN as a standardised test is the publication of schools’ results on the My School website, which became available in 2010 (ACARA, 2019a). Although this site presents contextual information about socio-economic background, a critical factor in student achievement, the NAPLAN test results are the main ‘numbers game’ for schools (Gannon, 2013, p19). The average score for a school and year level in each part of NAPLAN compared to the average for the nation can be displayed in graphs or tables. Nuances related to context tend to be lost in the face of such graphic wizardry.

That schools are being judged on the basis of performance in standardised tests is an indisputable fact. First and foremost, federal and state governments judge schools because they expect them to be accountable for the huge financial investment in education; the public as taxpayers are likely to share such expectations. Parents often use NAPLAN results to help them decide the ‘best’ school to send their children. Indeed, schools use NAPLAN results to judge themselves and each other, particularly in a free-enterprise climate in which schools are pitted against each other.

As a neat numerical summary of a school’s success in teaching literacy and numeracy, NAPLAN results are often used to make these judgements. Quantitative data allows for easy comparison between schools. People who are not educational professionals, including most politicians and parents, might latch on to such data because it does not require them to think too deeply about the complexity of school performance and the difficulties of effectively measuring it.

The media plays a key role as a mediator of NAPLAN data, interpreting and reframing the information and significantly influencing the judgements of all. Baroutsis (2016) claims that much media reportage is responsible for ‘negative, critical, oppressive and reductionist discourses’ (p 567) that encourage the perception that schools are in crisis. The media frequently use data to compare and rank schools, causing some to be ‘named, shamed and blamed’ (Elstad, 2009, p 173). Baroutsis cites evidence that media reports, often unfairly de-contextualising school performance, are contributing to the drift from government to non-government schools, thereby heightening the perception that state schools are underperforming, despite the fact that there are no significant differences in the performance of educational sectors (Goss et al, 2018b). Rather than framing the performance of schools by comparing NAPLAN results, Baroutsis argues that the media should draw on rich school-based narratives that contextualise school performance.

Baroutsis’ argument echoes the view of Gannon (2013) whose study focused on two schools with which she had personal experience. Gannon rejects the reductionist profile presented by the NAPLAN data and rather develops more situated accounts of teaching and learning at these schools. Such narratives, she argues, both augment and disrupt the ‘impoverished statistical stories’ (p 27) generated by the My School site.

Interestingly, the stakeholders most likely to be sceptical of NAPLAN data and its use to make judgements about schools are the educational professionals. A study of independent school principals and teachers found that only about a half considered NAPLAN results provided a useful diagnostic tool (Athanasou, 2010). In an analysis of submissions to the Senate Inquiry following the institution of NAPLAN/My School, the majority from educators, Ragusa and Bousfield (2017) concluded that overall there was an acceptance of NAPLAN as a data source but rejection of My School for contributing to ‘maladaptive practices’ (p 281), such as narrowing the curriculum and heightening teacher mistrust. Carter, Manuel and Dutton (2018) found that secondary English teachers generally questioned the purpose, validity and usefulness of NAPLAN literacy data and criticised the ways in which it encouraged teaching to the test, a tendency which Thompson and Harbaugh (2013) noted was particularly evident in low socio-economic status (SES) schools. Of course, primary teachers may have a more positive perception, given they are more likely to view teaching literacy and numeracy as core business in their work. Hardy (2014), for example, found that principals and teachers at three Queensland primary schools had appropriated the political pressure to improve NAPLAN results by using NAPLAN data for educative purposes.

No test is perfect – and NAPLAN is no exception. The NAPLAN tests assess student performance in four domains: reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. While the reading test is quite robust, using a variety of reading materials and comprehension questions, the writing test consists of just one task requiring either imaginative or persuasive writing. Unsurprisingly, schools have a tendency to focus on these two types of writing, often at the expense of other genres, in their writing programs (Spina, 2017). Language Conventions is perhaps the most problematic of the literacy tests, mostly using multiple-choice questions that require students to spot errors in brief, de-contextualised texts. These three literacy tests assess aspects of two communication modes, reading and writing, but ignore the literacy skills involved in the other four communication modes: speaking and listening, viewing and representing. Clearly, as a test of literacy, and as a basis therefore for making judgements about schools, the NAPLAN tests are limited and lack validity.

To the extent to which they attempt to compare students across the country, the NAPLAN tests also lack reliability. In particular, students sat the 2019 NAPLAN tests under different conditions: some students did the tests online, while others did paper tests; some students experienced frustrating technical glitches and had to complete the papers over two sessions, while others experienced no disruption. Changing the test conditions from year to year also raises questions about the reliability of comparisons over time.

Some researchers have taken issue with the validity and reliability of NAPLAN test results. Wu (2010) warns against making judgements about schools on the basis of NAPLAN data, claiming the margins of error for measuring student and school performance are too high. ‘It can be easily demonstrated that NAPLAN data do not provide the power to reflect school performance,’ she says. Caldwell (2010) criticises use of NAPLAN tests as a diagnostic tool because there is a five month gap between tests and schools receiving results. In fact, a former CEO of ACARA has said that NAPLAN tests should not be used as a diagnostic tool due to this time delay (Senate References Committee on Education, 2010). The practice of some schools of encouraging certain students not to attend tests in order to improve their results also raises questions about the validity of the tests (Australian College of Educators, 2010; Thompson et al, 2018).

Klenowski and Wyatt-Smith (2012) contend that a test, such as NAPLAN, at best provides a snapshot of performative achievement and cannot be relied upon to inform teaching and interventions for the improvement of learning. They argue instead to ‘bring teacher judgement to centre stage’ (p 75), to restore the value of school-based assessment: teachers using ongoing classroom assessment to identify students’ needs and tailor their teaching to address those needs. Their views accord with Lingard (2010), who says that standardised tests, such as NAPLAN, are not indicative of the impact of schooling and calls for recognition of teacher judgement in achieving improvements in student learning.

Alternative arguments supporting use of NAPLAN data to judge the performance of schools can be compelling and powerful. The Australian Curriculum, Assessment and Reporting Authority (ACARA), which manages the NAPLAN tests and My School website, provides an infographic on its website that identifies the benefits of NAPLAN (ACARA, 2019b - see Appendix 1). According to this source, parents get feedback on their children’s performance relative to other students across the country. Teachers and schools can use the data to ascertain how well their students are performing and adjust their teaching programs to suit students’ needs. For educational systems and governments, NAPLAN data shows how well students are achieving national standards and helps drive the agenda to improve school performance.

While there is some merit in these claims, one benefit cited in the ACARA infographic is particularly doubtful: ‘EQUITY: A fair go for all Australian students’. Presumably this means that NAPLAN data enables governments to allocate funding to ensure more equitable outcomes in education. According to a research publication on the Australian Parliament website (Clark, 2019), the Australian Government has increased overall funding to schools but the latest agreement ‘locks in underfunding’ for government schools, most of which will get 95% of the Schooling Resource Standard compared to 100% for more privileged non-government schools that already perform better in NAPLAN.

A conservative neoliberal think tank, the Centre for Independent Studies, endorses NAPLAN resoundingly. For example, a research report (Joseph, 2018) argues three main benefits of using NAPLAN data: teachers using the data for diagnostic purposes; transparency (available for all to see and discuss); and accountability (holding governments and schools accountable for the significant investment in education by parents and taxpayers). The same report dismisses arguments against NAPLAN. On the particular question of how schools might be harmed by judgements based on their NAPLAN results, Joseph claims that there have been no rigorous studies of how parents use My School data, and no substantial evidence to suggest My School has a negative effect on schools. Better, he says, that parents judge schools on their NAPLAN data rather than by school uniforms or information on school websites.

Peter Goss, director of the School Education Program at the Grattan Institute, is a keen advocate for NAPLAN. The Grattan Institute is a middle-of-the-road think tank on Australian public policy. In an online article for the ABC’s *The Conversation*, Goss (2018a) identifies all the benefits of NAPLAN mentioned in the ACARA infographic, but also argues that parents can make more informed choices about where to send their children to school. However, it is likely that many parents would unfairly compare and judge schools by their test results without considering the different contexts of those schools. He also argues that policy makers and researchers can use NAPLAN results to better understand student performance, thereby informing policies, such as the allocation of funding. However, political considerations often override data considerations in government policy.

While there is extensive research about the impact of standardised testing in contexts where it has been in operation for some time, especially USA and UK, we should be circumspect about such research because of important differences between the Australian NAPLAN/My School model and the USA and UK models. However, a thorough literature review by Polesel et al (2012) reveals Australian studies that suggest that NAPLAN promotes greater consistency, comparability and transferability of data across schools and jurisdictions (Santiago et al, 2011) and the potential for teachers to use the data for diagnostic purposes (Collier, 2010).

Drawing on my own experience as a Literacy Coordinator in a K-12 school (11 years) and Chief Education Officer, English and Literacy, for NSW Department of Education (8 years), I would agree that NAPLAN does have *potential* value as a tool to identify students’ strengths and weaknesses in literacy and numeracy; to help teachers tailor programs to address students’ needs; and to inform the appropriate allocation of resources for improving student learning outcomes. Realising that potential depends on the capacity of the tests to provide valid and reliable data and for all players to use that data (in addition to rich, school-based data) to make fair, well-founded judgements that lead to improved student learning outcomes. Interestingly, none of these potential benefits depend on public availability of NAPLAN data on the My School site, a situation that exposes schools to uneducated and unfair judgements, particularly by media and parents, that can affect them adversely.

The publication of NAPLAN results on the My School website is a practice that is not likely to cease in the short to medium term, given that it is supported by both sides of politics in Australia. However, the literature suggests that, overall, it is not reasonable, or indeed fair, to overrely on statistical data of dubious validity and reliability to judge the performance of schools. Educational leaders and professionals have a responsibility to take greater control of the agenda, to guide stakeholders to consider important contextual information in analysing NAPLAN data, to restore the pre-eminence of quality school-based assessment to monitor the performance of students and schools and to use school-based narratives to help all understand the complex issues faced by schools in improving student learning outcomes.

**References**

Athanasou, J. (2010). *NAPLAN and My School survey*. Barton, ACT: Independent Education Union of Australia.

Australian College of Educators (2010). Submission to the *Inquiry into the Administration and Reporting of NAPLAN Testing*, Senate References Committee on Education. Canberra: Employment & Workplace Relations.

Australian Curriculum, Assessment and Reporting Authority, ACARA (2019a). My School [website]. Sydney: ACARA.

Retrieved 13 September 2019 from:

https://www.myschool.edu.au

Australian Curriculum, Assessment and Reporting Authority, ACARA (2019b). NAPLAN provides benefits from the ground up for parents, schools and Australian education systems [infographic]. Sydney: ACARA.

Retrieved 8 September 2019 from:

[https://www.nap.edu.au/\_resources/Acara\_NAPLAN\_Infographic(V4-2).pdf](https://www.nap.edu.au/_resources/Acara_NAPLAN_Infographic%28V4-2%29.pdf)

Baroutsis, A. (2016). Media accounts of school performance: Reinforcing dominant practices of accountability. *Journal of Education Policy, 31*(5), 567-582.

Caldwell, B. J. (2010). How can we get off this bus? [paper presented at a symposium, *NAPLAN and the My School website: Issues and the way forward* at the 2010 International Education Research Conference of the Australian Association of Research in Education, University of Melbourne, 30 November 2010].

Carter, D., Manuel, J. & Dutton, J. (2018). How do secondary school English teachers score NAPLAN? A snapshot of English teachers' views. *Australian Journal of Language and Literacy, 41*(3), 144-154.

Clark, S. (2019). School funding and improving educational performance [briefing paper]. Canberra: Parliament of Australia.

Retrieved 8 September 2019 from: <https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook46p/SchoolFunding>

Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability 21*(2), 173–189.

Gannon, S. (2013). My school redux: Re-storying schooling with the My School website. *Discourse: Studies in the Cultural Politics of Education*, *34*(1), 17-30.

Goss, P. (2018a). Five things we wouldn’t know without NAPLAN. Melbourne: The Conversation.

Retrieved 8 September 2019 from: <https://www.abc.net.au/news/2018-05-15/naplan-testing-five-things-we-wouldnt-know/9761492>

Goss, P., Emslie, O. & Sonnemann, J. (2018b). Measuring student progress: A state-by-state report card [technical report]. Melbourne: Grattan Institute.

Retrieved 10 September 2019 from

<https://grattan.edu.au/wp-content/uploads/2018/10/910-Mapping-Student-Progress-Technical-Report.pdf>

Hardy, I. (2014). A logic of appropriation: Enacting national testing (NAPLAN)

in Australia. *Journal of Education Policy, 29*(1), 1-18.

Joseph, B. (2018). Why we need NAPLAN [research report]. Sydney: The Centre for Independent Studies.

Retrieved 8 September 2019 from: <https://www.cis.org.au/app/uploads/2018/05/rr36.pdf>

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.

Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education, 51,* 129–147.

Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The experience of education: The impacts of high stakes testing on school students and their families*. Sydney: The Whitlam Institute. Retrieved 6 September 2019 from:

[http://whitlam.org/\_\_data/assets/pdf\_file/0011/694199/The\_experience\_of\_ education\_-\_Qualitative\_Study.pdf](http://whitlam.org/__data/assets/pdf_file/0011/694199/The_experience_of_%20education_-_Qualitative_Study.pdf)

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 My School Senate Inquiry. *British Journal of Sociology of Education, 38*(3), 265-286.

Senate References Committee on Education, Employment & Workplace Relations (2010). Inquiry into the administration and reporting of NAPLAN testing. Canberra: Commonwealth of Australia.

Retrieved 7 September 2019 from: [http://www.aph.gov.au/senate/committee/eet\_ctte/naplan/report/ report.pdf](http://www.aph.gov.au/senate/committee/eet_ctte/naplan/report/%20report.pdf)

Spina, N. (2017). Governing by numbers: Local effects on students' experiences of writing, *English in Education*, *51*(1), 14-26.

Thompson, G., Adie, L. & Klenowski, V. (2018). Validity and participation: Implications for school comparison of Australia’s National Assessment Program, *Journal of Education Policy, 33*(6), 759-777.

Thompson, G. & Harbaugh, A. G. (2013). A preliminary analysis of teacher perceptions
of the effects of NAPLAN on pedagogy and curriculum. *Australian Educational Research, 40*, 299–314.

Wu, M. (2010). The (in)appropriate use of NAPLAN data*.* *Professional Voice, 8*(1), 21-25.

**Appendix 1:**

Infographic from the ACARA website (ACARA, 2019a): [https://www.nap.edu.au/\_resources/Acara\_NAPLAN\_Infographic(V4-2).pdf](https://www.nap.edu.au/_resources/Acara_NAPLAN_Infographic%28V4-2%29.pdf)

