



**Inquiry into the effectiveness of the National Assessment Program –
Literacy and Numeracy**

**Submission to the Education, Employment and Workplace Relations
References Committee**

Professor Brenton Doecke
Dr Brian Doig
Associate Professor Alex Kostogriz
Professor Christine Ure

Submitted on Behalf of the School of Education
Deakin University

The National Assessment Program – Literacy and Numeracy (NAPLAN) is one of a number of assessment programs currently implemented by the Australian Federal Government in collaboration with state and territory governments. The declared aim of these programs is to enable various stakeholders in education, most notably parents, ‘to monitor student progress over time against national and international standards’. This is in order to support ‘student learning by providing schools with information that enables strengths and weaknesses in teaching programs to be identified, thereby improving the learning outcomes for all Australian students’ (http://www.curriculum.edu.au/verve/resources/NAP_2009-Assess_Stud_Achiev_Aust-Parent_Info_Brochure.pdf)

NAPLAN is arguably the most controversial of these assessment programs for reasons that we shall outline in this submission. Since its inception in 2008, debate has continued amongst academics, teachers and the wider community about its impact on Australian education and the quality of teaching and learning in Australian schools. Only recently NAPLAN has been the subject of debate in the *Melbourne Age*, which has featured several articles by teachers and other experts about its impact on schools. The key criticism is that NAPLAN has led to a narrowing down of the curriculum, with some schools spending an inordinate amount of time preparing their students to sit for the test. A related criticism is that the publication of results on the *My School* website has produced a de facto league table of schools that unfairly represents a school’s performance and puts further pressures on schools to focus on NAPLAN at the expense of other areas of the curriculum.

The following submission does not engage with all the aspects of NAPLAN. It is written from the perspective of committed teacher educators and researchers within the field of education at Deakin University, and draws on both their research and their experience of the impact that NAPLAN has had on the professional learning of student teachers within their teacher education courses.

The Perspective of Teacher Educators at Deakin University

During the past five years, Professor Doecke and Associate Professor Alex Kostogriz have been Chief Investigators in a major research project funded by the Australian Research Council that examined the impact of NAPLAN on teachers' work, focusing specifically on the work of English literacy teachers in primary and secondary schools. Some of the comments that follow derive from this project. Professor Ure adds a perspective from the MCEEDCYA review of student engagement from Australian schools with a low Index of Community Socio-Economic advantage (ICSEA) reported in 2012. This submission also draws on the expertise of Dr Brian Doig, who has extensive experience in the field of educational testing. In addition, we make use of comparative analyses recently conducted by Associate Professor Kostogriz on the role that standardized testing plays in a diverse range of policy settings around the world. Another important source of evidence for the claims that follow is the experience of teacher educators within the Faculty who teach within pre-service programs, many of whom have concerns about the disruptive effect that NAPLAN has had on the professional learning of pre-service students when they complete their teaching rounds during the time when NAPLAN tests are being implemented.

Given the nature of our expertise, we feel that we are best placed to address the following items in the terms of reference:

- a) whether the evidence suggests that NAPLAN is achieving its stated objectives,
- b) unintended consequences of NAPLAN's introduction,
- d) the impact on teaching and student learning practices of publishing NAPLAN results on the MySchool website
- e) potential improvements to the program, to improve student learning and assessment and
- f) international best practice for standardized testing, and international case studies about the introduction of standardized testing.

Rather than simply addressing each item separately, however, we feel that it is important to begin by identifying the key problem posed by NAPLAN,

which relates to its conflicting objectives as a diagnostic tool of students' literacy and numeracy abilities and its role in enforcing school accountability vis-à-vis the general public. This addresses items a, b, and d.

The Problem in a nutshell (a, b & d)

The issue is not whether standardized testing can serve a useful educational function. Standardised testing has been around for a lot longer than NAPLAN, and most (if not all) teachers have used standardised tests at some stage in their careers, whether for diagnostic purposes or simply to gain a snapshot of the range of abilities of the students in their classes. The question is whether such tests can meaningfully be used to represent the quality of the teaching and learning that occur at a whole school level, especially when they take the form of summative judgments (as on the *MySchool* website).

Several of the teachers interviewed as part of the ARC project mentioned above repeatedly questioned the diagnostic value of tests, when the results only become available several months after the test has been done. They felt that the test provided only a crude indication of a student's ability in comparison with their day-to-day assessment of that student's work. As a result, they saw the test as undermining their authority as teachers, specifically with respect to their capacity to make informed, professional judgments about the learning needs of the young people in their classrooms. If the test has validity as a representation of student's ability, it remains valid only for the time the test was conducted. Students typically move on, especially in the time between taking the test and learning the results, and so the test is of limited value when it comes to addressing the literacy and numeracy needs of individual students.

But the key issue, as we have indicated, is not the validity of the test as such. There are undoubtedly refinements that could be made to NAPLAN that would enhance its capacity to diagnose the literacy and numeracy needs of individual students and to identify areas where curriculum and pedagogy could be improved (some suggestions follow). This goal, however, is being compromised by the publication of results on the *MySchool* website, which reduces the test results to summative judgments

of a whole school's performance, sometimes unfairly stigmatizing the school in the eyes of a general public that does not fully appreciate the complexities of teaching and learning and the complex judgments involved in assessing students' abilities, especially with respect to addressing the needs of culturally diverse and disadvantaged communities. Rather than transparency and accountability, the *MySchool* website is blind to the full range of educational achievement and the professional responsibility that school leaders and teachers take in their efforts to enhance the educational experience of their students. NAPLAN might be reconceptualised in order to more effectively support the learning needs of all students in Australian schools, but this cannot be done as long as it is used as a vehicle to impose a form of accountability that puts schools on a competitive basis with each other, as though its primary purpose is as a vehicle to exercise consumer choice (cf. Koretz, 2008).

International best practice in standardised testing (f)

Concerns about teaching quality and teacher accountability are reflected in the educational policy-making of high performing and rapidly improving jurisdictions in literacy and numeracy, as measured by PISA, TIMMS and PIRLS. However, the ways of improving the quality of teaching, curriculum standards and accountability in these jurisdictions are significantly different from that represented by NAPLAN.

Unfortunately, the introduction of NAPLAN in Australia has followed the logic of accountability policies in such countries as the USA and the UK. In these countries, high-stakes tests, such as literacy and numeracy assessment programs, assess the general capabilities and skills of students and have consequences for students in terms of intervention, or remedial measures that may need to be taken, and for teachers in terms of punitive outcomes. Research on high-stakes testing has found that they may have a negative impact on teachers with a resultant degradation of students' experiences of learning. The impact of this may be defined as a shift from a focus on the needs of the child to the needs of the evaluation and reporting process (Armein & Berliner, 2002; Au 2007, 2008).

The analysis of high performing jurisdictions conducted by OECD (2012a,

2012b) demonstrates either a shift from high-stakes testing to *school-based* forms of assessment or to using test results to guide intervention based on the professional judgement of teachers. An example of such a jurisdiction is Ontario (Canada) where reforms have shifted the focus away from punitive accountability measures and performance pay and towards building shared sociocultural and leadership values and purposes within a system that aspires to improve (Fullan, 2006). The key to this strategy is a professionally-driven rather than market-driven system change. Levin (2007) describes this approach as follows:

The Province of Ontario's education change strategy embodies vital principles, grounded in research, that are associated with meaningful and sustainable change. Changes are respectful of professional knowledge and practice. Main elements of change are coherent and aligned at the provincial, district and school level. Key partners – the provincial Ministry of Education, school boards, schools, and provincial and local organizations of teachers, principals, and other partners – work together. Change strategies are comprehensive and emphasise professional learning, strong leadership, necessary resources, and effective engagement of parents and the broader community. We believe that this strategy provides an example of large-scale change that is effective and sustainable.

On the basis of this professionally-driven approach, Ontario balances administrative and professional accountability and has seen the dramatic reduction of low-performing schools (Levin & Fullan, 2008). The underlying assumption of Ontario's leaders seems to be that teachers are professionals who can be trusted and are generally motivated but require on-going support and investment in their capacity building. Consequently, teachers seem to take more responsibility for their own performance and students' learning. This is often not the case in countries with a more punitive approach to external accountability.

Similar developments have occurred in Hong Kong, Shanghai and Singapore where accountability is not perceived as an external add-on but is rather built into the system of expectations and relations, as well is an essential part of teacher professionalism.

An example from Australian schools highlights the importance of these characteristics to the success of students from areas with a low Index of Community Socio-Economic Advantage. Improvements in student engagement were a consequence of the capacity of schools to address the multiple social and learning needs of students in their community (Ure and Gray, 2012). High quality leadership, non-judgemental professional collaboration among teachers and the extensive use of data to make regular assessments of, and adjustments to, student learning characterised successful schools. Teachers sought reliable and timely information to assist them to make changes in student learning groupings and instructional techniques. Their professional accountability was vested in meeting the day to day learning needs of their students and being able to track outcomes using reliable sources of data.

Finland, one of the top countries in literacy and numeracy outcomes, has not followed a trend of using high-stakes testing to increase teacher accountability and raise student performance. Finnish policy makers realized early on that quality teaching, rather than externally set standards and standardized testing, is the key element that makes a difference in what students learn in school. Teachers' work is driven therefore by their responsibility for curriculum, teaching and learning, rather than testing. Finns have adopted 'smart' accountability policies that embed student assessment in the teaching and learning process. Assessment is used to improve both teachers' and students' work, grounding assessment practice in the national curriculum that prioritises individualized education and creativity. As a result, students are judged more against their individual progress and abilities rather than against statistical data. Since Finnish teachers must design and conduct appropriate curriculum-based assessments to document student progress, classroom assessment and school-based evaluation are important parts of teacher education and professional development.

In sum, high performing jurisdictions attempt to avoid the disadvantages often associated with external high-stakes testing, such as narrowing of the curriculum, teaching to the test (narrowly conceived), unhealthy competition among schools and negative effects of the testing regime on

the well-being of teachers and students. In doing so, policy-makers see the value of assessment primarily in terms of its capacity to positively affect student learning, rather than a means of increasing student scores on a particular test and holding teachers accountable for this. If student learning remains unaffected, or if testing leads to biased teaching, the validity of such high-stakes tests is questioned. The teaching workforce in high performing jurisdictions, and particularly in Finland, is not convinced that external high-stakes testing and accountability built on test results are beneficial to students and their learning.

How might NAPLAN be improved? (e)

Some of the criticism of NAPLAN has been directed towards its technical aspects, such as the accuracy of data representation on the *MySchool* website (Wu, 2010). Our focus here in this final section of our submission is the assessment instrumentation and reporting methods, which are open both to critique and possible improvement.

Teaching to the test

Firstly, teaching to the test is not necessarily a bad thing, if the test represents the curriculum to be studied. NAPLAN is not completely unanchored in the curriculum. The issue is the small number of items used to represent the curriculum and to make generalisations about the quality of teaching and learning in individual schools. The ‘cherry-picking’ of curriculum topics to be the NAPLAN content leads to some disquieting consequences, such as teachers trying to guess exactly what to concentrate upon in their classes.

How could this flaw be eradicated? Why not have more items in the tests? Obviously the number of items presented to students in a given time is a limiting factor, and yet, in large-scale international assessment programmes like *Trends in Mathematics and Science Study* (TIMSS) and the *Programme for International Student Assessment* (PISA), over 200 items might be used to represent the curriculum. How is this achieved without imposing an extraordinarily long test? The issue is solved by the use of several test booklets, each with a small number of items from the total, large pool of

items. In this way, large-scale assessment programmes are able to gain better data on a broader range of curriculum topics, and provide more useful information about student learning if they wish to do so (Doig, 2006).

The linking of the smaller test booklets, to form a quasi-whole, is easily done by the use of a small core of common items. These common items appear in all booklets, and the data analysis procedures, used to analyze current NAPLAN tests, accommodates such a data gathering structure. This arrangement has been in use internationally at least since the mid-1990s. Further, the Australian Government accepts results from TIMSS and PISA without complaint that students responded to booklets with a common core of items and the inclusion of different items. But for NAPLAN, no-one appears to wish to contemplate this as a solution to a significant flaw in the current assessment programme.

A more serious flaw in the current programme is the poor level of reporting to stakeholders. In particular, classroom teachers are poorly served by the reporting of their students' performance, despite the many opportunities for better reporting. In a paper examining the reporting of large-scale assessment programmes, Doig (2006) described different reporting formats as being at varying 'distances' from being useful in the classroom: that is, how the reports inform teachers of the strengths and weaknesses of their students. NAPLAN reporting essentially *counts* student errors (*summative* assessment), but what teachers want, and need, are reports that *account* for student errors (*formative* assessment) (Black & Wiliam, 1998; Black, Harrison, *et al.*, 2002).

Formative information, from NAPLAN data, requires a simple extension of the NAPLAN item design, and can build on the current form of data analysis. The end result could be a simple formative report that any teacher could use to diagnose their students' failings in NAPLAN assessments. How can this be done? Doig (2012) demonstrated, using TIMSS data, that a slight additional analysis to the programme would provide, through simple graphical means, a report that gave teachers very specific, formative, information from the results of this large-scale assessment. That is, for every response to an item, a direct link can be made between students'

overall performance and their likely response. The cost of such an improvement to the NAPLAN reporting would be very small in relation to the overall cost of the programme, and this change would benefit, not only countless students, but also remove many of the criticisms of the NAPLAN programme.

Concluding Comments

We have raised questions with respect to the generalizability of NAPLAN's findings, most notably its capacity to adequately represent the quality of the teaching and learning that occurs within any particular school. To improve the test would mean giving much stronger emphasis to refining its capacity to provide support for teachers in their efforts to enhance the educational opportunities of young people in their classrooms (i.e. as an instrument for formative assessment). However, this cannot be done as long as the test is used for the purposes of school accountability that puts schools on a competitive footing, as though it is a question of deciding which school is doing the 'best' (when the test is effectively used as a form of summative assessment).

The unintended consequences of NAPLAN largely arise from this confusion of purposes.

In closing, we wish to draw attention to another unforeseen consequence of imposing this type of testing on the nation's schools, namely its effect on student teachers' learning during the practicum. NAPLAN week is not a good time for our student teachers to be doing their practicum: rather than being able to implement curriculum and to interact with young people in classroom settings, they are typically reduced to observing the routines associated with the administration of NAPLAN tests. They return to university with stories about how schools have been obliged to completely reorganise their timetables and classrooms in order to administer the tests, as well as insights into the way teachers and their students have coped with testing on this scale. This constitutes a form of professional learning for them, but it is a poor substitute for the kind of learning they would normally experience if they were to visit schools at another time of the year.

For all the claims made that NAPLAN does not lead to major disruption of the everyday life of schools, the reports of our student teachers convey a very different impression. NAPLAN week is not a richly rewarding professional experience for them, and we suspect that in schools themselves worthwhile curriculum and pedagogy are likewise put on hold, until the tests are completed.

References

- Armein, A. & Berliner, D. (2002). High stakes testing, uncertainty and student learning, *Education Policy Analysis Archives*, 10 (18), accessed 29 May 2013, <http://epaa.asu.edu/epaa/v10n18/>
- Au, W. (2007). High-stakes testing and curricular control: a qualitative meta-synthesis. *Educational Researcher*, 36(5), pp. 258–267.
- Au, W. (2008). Between education and the economy: high-stakes testing and the contradictory location of the new middle class. *Journal of Education Policy*, 23 (5), pp. 501-513.
- Australian Curriculum and Assessment Authority (ACARA) (2013) The Australian Curriculum. Retrieved from: <http://www.australiancurriculum.edu.au/>
- Black, P., Harrison, C., Lee, C., Marshall, B, & Wiliam, D. (2002). Working inside the black box: Assessment for learning in the classroom. London: Department of Education and Professional Studies, King's College.
- Black, P., & Wiliam, D (1998). Assessment and classroom learning. *Assessment in Education: principles, policy and practice*, 5(1), 7-74.
- Doig, B. (2006). Large-scale mathematics assessment: looking globally to act locally. *Assessment in education: principles, policy and practice*, 13(3), 265-288.
- Doig, B. (2012). *Reporting formative and summative information on a single scale*. Unpublished Research Dissertation, Deakin University, Melbourne.
- Fullan, M. (2010). *All Systems Go: The Imperative for Whole System Reform*. Thousand Oaks, CA: Corwin Press
- Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, Massachusetts: Harvarad University Press.

- Levin, B. (2007). A strategy for large-scale improvement in education: The Ontario approach. Paper presented for the Australian Curriculum Conference.
- Levin, B., & Fullan, M. (2008). Learning About System Renewal. *Educational Management Administration Leadership*, 36(2), 289-303.
- OECD (2012a). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing.
<http://dx.doi.org/10.1787/9789264130852-en>
- OECD (2012b). *Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education*, OECD Publishing.
- Ure, C., & Gray, J. (2012) Research and Mapping Project for MCEECDYA, Student Academic Engagement, Ministerial Committee for Education, Early Childhood Development and Youth Affairs.
- Wu, Margaret (2010) *The Inappropriate Use of NAPLAN Data*. Professional Voice, 8 (1). pp. 21-25. and from:
http://www.aeuvic.asn.au/pv_8_1_complete.pdf