

Mathematics for All? The Case for and Against National Testing

Gilah C. Leder

Abstract National numeracy tests were introduced in Australia in 2008. Their format and scope are described and appraised in this paper. Of the various group performance trends presented in the annual national NAPLAN reports two (gender and Indigeneity) are discussed in some detail. For these, the NAPLAN findings are compared with broader international data. Recent Australian research spawned by, or benefitting from, the NAPLAN tests is also summarised. In some of this work, ways of using national test results productively and constructively are depicted.

Keywords National tests • Gender • Indigeneity

Introduction

It should come as no surprise... that the introduction of a national regime of standardised external testing would become a lightning rod of claim and counter-claim and a battleground for competing educational philosophies. The National Assessment Program—Literacy and Numeracy (NAPLAN) is a substantial educational reform. Its introduction has been a source of debate and argument (Sidoti and Keating 2012, p. 3).

Formal assessment of achievement has a long history. Kenney and Schloemer (2001) point to the use, more than three thousand years ago, of official written examinations for selecting civil servants in China. The birth of educational assessment is, however, generally traced to the 19th century and its subsequent growth has undoubtedly been intertwined with advancements in the measurement of human talents and abilities (Lundgren 2011). Over time the development of large scale, high stake testing and explorations of its results have proliferated. “Many nations”, wrote Postlethwaite and Kellaghan (2009), “have now established national assessment mechanisms with the aim of monitoring and evaluating the quality of their education systems across several time points” (p. 9). More recently, Eurydice (2011) also drew attention to the widespread practice of national testing

G.C. Leder (✉)

Monash University and La Trobe University, Melbourne, Australia
e-mail: Gilah.Leder@monash.edu; G.Leder@latrobe.edu.au

throughout Europe, confined in some countries to a limited number of core curriculum subjects but in others comprising a broad testing regime. Large scale national assessment programs, with particular emphasis on numeracy and literacy¹, were introduced in Australia in 2008—after extensive consultation and much heated debate within and beyond educational and political circles.

The NAPLAN Numeracy Tests

Until 2007, Australian states and territories ran their own numeracy and literacy testing programs. Although much overlap could be found in the assessment instruments used in the different states, there were also variations—some subtle, others substantial—in these tests.

The first National Assessment Program—Literacy and Numeracy (NAPLAN) tests were administered in May 2008 and have been conducted annually since then. For the first time, students in Years 3, 5, 7, and 9, irrespective of their geographic location in Australia, sat for a common set of tests, administered nation-wide. The Numeracy tests contain both multiple choice and open-ended items. Their scope and content are informed by the *Statements of Learning for Mathematics* (Curriculum Corporation 2006). The ‘what’ students are taught is described by four broad numeracy strands. These are Algebra, function and pattern; Measurement, chance and data; Number; and Space, though some questions may overlap into more than one strand. Instructional strategy, the ‘how’ of mathematics is described by proficiency strands. “The proficiency strands—Understanding, Fluency, Problem solving and Reasoning—describe the way content is explored or developed through the ‘thinking’ and ‘doing’ of mathematics” (Australian Curriculum, Reporting and Assessment Authority ACARA 2010). In Years 3 and 5, the papers are expected to be completed without calculator use. Two distinct papers are set for Year 7 and 9 students—one is expected to be completed without the use of a calculator; for the other calculator usage is allowed.

The NAPLAN numeracy scores for Years 3, 5, 7, and 9 are reported on a common scale which is divided into achievement bands. For each of these year levels, the proportion of students with scores in the six proficiency bands considered appropriate for that level is shown. For Year 3, 5, 7, and 9 these are bands one to six; three to eight; bands four to nine; and bands five to ten respectively. Each year, results of the NAPLAN tests are published in considerable detail, distributed to each school, and made readily available to the public.

The advantages anticipated by the introduction of national tests to replace the variety of tests previously administered by the different Australian states and

¹ Sample assessment tests have been administered to selected groups of students in Years 6 and 10 in Scientific Literacy (Year 6 students only), Civics and Citizenship, and Information Communication Technology Literacy. These sample assessments were introduced respectively in 2003, 2004, 2005 and are held on rolling a three-yearly basis.

territories were similar to those commonly put forward in the wider literature (e.g., Postlethwaite and Kellaghan, 2009) as a rationale or justification for introducing national tests: assessment consistency across different constituencies, increased accountability, and a general driver for improvement.

ACARA is responsible for the development of the national assessment program and the collection, analysis, and reporting of data. The procedures followed are described clearly on the ACARA website and are consistent with those generally advocated for large scale assessment testings (Joint committee on testing practices 2004). Guidance on interpreting the vast amount of data in the National Report is provided in the document itself (ACARA, 2011a) and in multiple ancillary documents (see e.g., ACARA, 2011b; Northern Territory Government n.d). NAPLAN achievement outcomes are reported not only at the national level, but also by state and territory data; by gender; by Indigenous status; by language background status²; by geolocation (metropolitan, provincial, remote and very remote); and by parental educational background and parental occupation. Each of these categories which are clearly not mutually exclusive, has been shown, separately, to have an impact on students' NAPLAN score. Broad performance trends for the different groupings have been summarised as follows:

In Australia, girls have typically performed better on tests of verbal skills..., while boys have typically performed better on tests of numerical skills... Children from remote areas, children from lower socioeconomic backgrounds and children of Indigenous background have tended to perform less well on measures of educational achievement (NAPLAN 2011b, p. 255).

It is beyond the scope of this paper to look at each of the categories mentioned above. Instead, the focus is on two groups of special interest: *girls/boys* and *Indigenous* students. What trends can be discerned in the years of NAPLAN data available at the time of writing this paper?

Trends in NAPLAN Data: Gender and Indigeneity

Data for Years 3 and 9 by gender and Indigeneity are shown in Tables 1 and 2 respectively.

From these tables it can be seen that:

Gender

- The mean NAPLAN score for males is invariably higher than that for females.
- The standard deviation for males is also consistently higher than for females, that is the range of the NAPLAN scores for males is higher than that for females.

² LBOTE, language background other than English, defined as “A student is classified as LBOTE if either the student or parents/guardians speak a language other than English at home.”

Table 1 Numeracy Year 3 students, NAPLAN achievement data 2008–2011

Group \year		All	M	F	Indigenous	Non-Indigenous	Indigenous year 5 ^a
2008	Mean	396.9	400.6	393.1	327.6	400.5	408.0
	S.D	70.4	72.8	67.6	70.6	68.4	65.8
	≥National min ^b (%)	95.0 %	94.6 %	95.5 %	78.6 %	96.0 %	69.2 %
2009	Mean	393.9	397.5	390.2	320.5	397.7	420.5
	S.D	72.9	75.3	70.0	76.0	70.6	66.4
	≥National min (%)	94.0 %	93.5 %	94.5 %	74.0 %	95.2 %	74.2 %
2010	Mean	395.4	397.8	392.9	325.3	399.0	416.9
	S.D.	71.8	74.0	69.3	71.2	69.8	70.5
	≥National min (%)	94.3 %	93.7 %	94.9 %	76.6 %	95.3 %	71.4 %
2011	Mean	398.1	402.6	393.5	334.4	401.7	421.1
	S.D.	70.6	73.0	67.6	65.0	69.1	64.0
	≥National min (%)	95.6 %	95.2 %	96.0	83.6 %	96.4 %	75.2 %

^a I refer to the data in the last column later in the paper. To save space the information is included in this table

^b National minimum standards: The second lowest band on the achievement scale represents the national minimum standard expected of students at each year level

Table 2 Numeracy Year 9 students, NAPLAN achievement data 2008–2011

Group \year		All	M	F	Indigenous	Non-Indigenous	Year 7 Non-Indigenous
2008	Mean	582.2	586.5	577.6	515.1	585.7	548.6
	S.D	70.2	72.0	68.1	65.6	68.7	71.6
	≥National min (%)	93.6 %	93.7 %	93.6 %	72.5 %	94.8 %	96.4 %
2009	Mean	589.1	592.4	585.6	520.2	592.4	547.0
	S.D	67.0	69.2	64.4	63.2	65.3	69.4
	≥National min (%)	95.0 %	94.7 %	95.2 %	75 %	96 %	95.8 %
2010	Mean	585.1	591.1	578.8	515.2	588.5	551.4
	S.D	70.4	72.7	67.4	64.7	68.8	70.8
	≥National min (%)	93.1 %	93.3 %	92.9 %	70.4 %	94.3 %	96.1 %
2011	Mean	583.4	589.3	577.3	515.8	586.7	548.5
	S.D	72.1	74.7	68.7	62.2	70.8	72.1
	≥National min (%)	93.0 %	93.0 %	93.0 %	72 %	94.1 %	95.5 %

(Data in both tables adapted from ACARA 2011a)

- At the Year 3 level a higher proportion of females than males score above the national minimum standard NAPLAN score. There is no such consistency at the Year 9 level, with a marginally higher proportion of males performing at or above the minimum level in some years (e.g., 2008, 2010) and a marginally higher proportion of females performing at or above the minimum level in other years (e.g., 2009).

Indigeneity

- Each year, non-Indigenous students do (a lot) better than Indigenous students. From Table 1 it can be seen that Year 5 Indigenous students performed just above the level of Year 3 non-Indigenous students; from Table 2 that Year 9 Indigenous students performed below the level of Year 7 non-Indigenous students.
- In 2011, there was a noticeable increase, compared with the previous years, in the percentage of Indigenous students at Year 3 who performed at or above the national minimum standard. No such increase is apparent at the other Year levels.

Also relevant are the following:

- In 2011, between 240,000 and 250,000 non-Indigenous students sat for the Years 3, 5, 7, and 9 NAPLAN papers. For the Years 3, 5, and 7 papers close to 13,000 Indigenous students participated. A smaller number, about 10,000 sat for the Year 9 paper. Thus at the different Year levels, Indigenous students comprised between 4 and 5 % of the national groups involved in the NAPLAN tests.³
- The exemption rates for the two groups are similar: around 2 % for Indigenous students and about 1 % for non-Indigenous students.

These summaries for gender and Indigenous performance outcomes are set against a broader context in the next sections.

Gender

In many countries, including Australia, active concern about gender differences in achievement and participation in mathematics can be traced back to the 1970s. Two reliable findings were given particular prominence: that consistent between-gender differences were invariably dwarfed by much larger within-group differences; and that students who opted out of post compulsory mathematics courses often restricted their longer term educational and career opportunities. These generalizations remain relevant.

³ The proportion of school students in Australia identified as Aboriginal and/or Torres Strait Islanders has risen from 3.5 % in 2001 to almost 5 % in 2011(<http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4221.0main+features402011>).

Evidence of progress towards gender equity more broadly than with respect to mathematics learning specifically has been mapped in many different ways:

Whereas the challenge of gender equality was once seen as a simple matter of increasing female enrolments, the situation is now more nuanced, and every country, developed and developing alike, faces policy issues relating to gender equality. Girls continue to face discrimination in access to primary education in some countries, and the female edge in tertiary enrolment up through the master's level disappears when it comes to PhDs and careers in research. On the other hand, once girls gain access to education their levels of persistence and attainment often surpass those of males. High repetition and dropout rates among males are significant problems (UNESCO 2012, p. 107).

As can be seen from large scale data bases such as NAPLAN, some gender differences in mathematics performance remain. What explanations for this have been proffered?

Explanatory Models

Over the years a host of, often subtly different, explanatory models for gender differences in mathematics learning outcomes have been proposed. They invariably contain a range of interacting factors—both person-related and environmental. Common to many models is an

...emphasis on the social environment, the influence of other significant people in that environment, students' reactions to the cultural and more immediate context in which learning takes place, the cultural and personal values placed on that learning and the inclusion of learner-related affective, as well as cognitive, variables (Leder 1992, p. 609).

A comprehensive overview of research concerned with gender differences in mathematics learning is beyond the scope of this paper. Instead, some recent publications, the majority with at least a partial cross-national perspective and published in a variety of outlets, are listed to sketch the range of factors invoked as explanatory or contributing factors for the differences still captured. Included is work in which the need for a repositioning of perspective to examine gender differences, via a different theoretical (often feminist and/or socio-cultural) framework, is prosecuted, as well as several articles in which there are strong attempts to rebut the notion that gender differences persist.

Gender Differences: Possible Explanations

- Kaiser et al. (2012) found, in a large study involving over 1,200 students, that “the perception of mathematics as a male domain is still prevalent among German students, and that this perception is stronger among older students. This is either reinforced by the peer group, parents or teachers” (p. 137).
- Kane and Mertz (2012) concluded “that gender equity and other sociocultural factors, not national income, school type, or religion per se, are the primary determinants of mathematics performance at all levels of boys and girls” (p. 19).

- Stoet and Geary (2012) challenged but ultimately supported the notion of stereotype threat (provided it is carefully operationalized) as an explanation for the higher performance of males in mathematics, particularly at the upper end.
- Wai et al. (2010) examined 30 years of research “on sex differences in cognitive abilities” and focussed particularly on differences in favour of males found in the top 5 %. As well as highlighting the role of sociocultural factors they concluded: “Our findings are likely best explained via frameworks that examine multiple perspectives simultaneously” (p. 8).
- “Traditionally, all societies have given preference to males over females when it comes to educational opportunity, and disparities in educational attainment and literacy rates today reflect patterns which have been shaped by the social and education policies and practices of the past. As a result, virtually all countries face gender disparities of some sort” (UNESCO 2012, p. 21).

Gender Differences: Have They Disappeared?

- Else-Quest (2010) used a meta-analysis of PISA and TIMSS data to examine the efficacy of the gender stratification hypothesis (that is, societal stratification and inequality of opportunity based on gender) as an explanation for the continuing gender gap in mathematics achievement reported in some, but not in other, countries. They concluded that “considerable cross-national variability in the gender gap can be explained by important national characteristics reflecting the status and welfare of women” (p. 125) and that “the magnitude of gender differences in math also depends, in part, upon the quality of the assessment of mathematics achievement” (p. 125).
- Hyde and Mertz (2009) drew on contemporary data from within and beyond the U.S. to explore three major questions: (1) “Do gender differences in mathematics performance exist in the general population? (2) Do gender differences exist among the mathematically talented? (3) Do females exist who possess profound mathematical talent?” (p. 8801). They summarised respectively: (1) Yes, in the U.S. and also in some other countries; (2) Yes, there are more males than females among the highest scoring students, but not consistently in all ethnic groups. Where this occurs, the higher proportion of males is “largely an artefact of changeable sociocultural factors, not (due to) immutable, innate biological differences between the sexes” (p. 8801); and (3) Yes, there are females with profound mathematical talent.

Gender Differences: Looking for New Directions

- Erchick (2012) argued that consideration of conceptual *clusters*, rather than topics in relative isolation, should lead to new questions in as yet fallow ground to be found in the field of gender differences in mathematics. Three clusters are

proposed: “Feminism/Gender/Connected Social Constructs; Mathematics/Equity/Social Justice Pedagogies; and Instruction/Perspectives on Mathematics/Testing” (p. 10).

- Jacobsen (2012) is among many of those who argue for a reframing of the deficit model approach to gender differences in which male performance and experience are considered the norm to one recognizing the social construction of gender and accepting that females may learn in different, but not inferior, ways from males. One approach to translating this theoretical perspective into practice is also described.

In some of the publications listed (as well as in others not listed here) gender differences are minimized while in others they are given centre-stage. Collectively, a complex rather than simplistic network of interweaving and sometimes contrasting pressures emerges from this body of work. After four decades of research on gender and mathematics, there is only limited consensus on the size and direction of gender differences in performance in mathematics and stark variation in the explanations put forward to account when differences are found.

The NAPLAN scores summarised in Tables 1 and 2 also require a nuanced rather than uni-dimensional reading. When performance on the NAPLAN test is described in terms of mean scores, the small but consistent gender differences in favour of males mirror those obtained in other large scale tests such as the Trends in International Mathematics and Science Study (TIMSS) and the OECD Programme for International Student Assessment (PISA)⁴. But in terms of another set of NAPLAN achievement criteria, the percentage of students achieving above the minimum national average, the small differences reported generally favour girls in the earlier years of schooling, in each of 2008–2011 at Year 3; for three of the four years (2009–2011) for Years 5 and 7; but in only one year (2009) at the Year 9 level. Clearly, gender differences in performance on the NAPLAN tests are small, consistent or variable, depending on the measuring scale and the method of reporting used.

Assessment: Gender Neutral or not?

That gender differences in mathematics learning may be concealed or revealed by the assessment method used is not a new discovery. Else-Quest et al. (2010) judged that “the magnitude of gender differences in math also depends, in part, upon the quality of the assessment of mathematics achievement” (p. 125). Dowling and

⁴ Differences in the samples involved in the three tests are worth noting. NAPLAN is administered to all students in Years 3, 5, 7, and 9. It is best described as a census test. The TIMSS tests, aimed at students in Years 4 and 8, and the PISA tests administered to 15-year-old students, are restricted to “a light sample (of) about 5 % of all Australian students at each year or age level” (Thomson, p. 76).

Burke (2012) pointed to the 2009 General Certificate of Secondary Education examinations in the U.K. as the first occasion in a decade for boys to perform better than girls in an external examination. “This reversal coincided with a change in the form of the examination” (p. 94), they noted.

A now somewhat dated, yet still striking, example of the impact of the format of examinations on apparent gender differences in mathematics achievement is provided by Cox et al. (2004). They tracked gender differences in performance in the high stake, end of Year 12 examinations in Victoria, Australia for the years 1994–1999, a sustained period of stability in the state’s external assessment regime. Student performance in three different mathematics subjects—Further Mathematics (the easiest and most popular of the three mathematics subjects offered at Year 12), Mathematical Methods (a pre-requisite for many tertiary courses), and Specialist Mathematics (the most demanding of the three mathematics subjects)—were among the results inspected. For each of these three subjects there were three different examination components. These were common assessment task (CAT) 1 consisting of a school assessed investigative project or problem, to be completed over several weeks; CAT 2, a strictly timed examination comprising multiple choice and short answer questions; and CAT 3, also a strictly timed examination paper with problems requiring extended answers. Thus CATs 2 and 3 followed the format of traditional timed examinations.

During the period monitored, a student enrolled in a mathematics subject in Year 12 was required to complete three assessment tasks in that subject. A test of general ability was also administered to the Year 12 cohort. These combined requirements provided a unique opportunity to compare the performance of the same group of students on timed and untimed examinations and on papers with items requiring substantially and substantively different responses. In brief:

- Males invariably performed better (had a higher mean score) than females on the mathematics/science/technology component of the general ability test.
- In Further Mathematics, females outperformed males in CAT 1 and in CAT 2 in all of the six years of data considered, and on CAT 3 for five of the six years.
- In Mathematical Methods, females performed better than males in all of the six years on CAT 1; males outperformed females on CAT 2 and CAT 3 for the six years examined.
- In Specialist Mathematics, females performed better than males in all of the six years on CAT 1 and in five of the six years on CAT 3. However males outperformed females on CAT 2 for each of the six years examined.

Thus whether as a group males or females could be considered to be “better” at mathematics depends on which subject or which test component is highlighted. If the least challenging and most popular mathematics subject, Further Mathematics, is referenced then the answer is females. If for all three mathematics subjects the focus is confined to the CAT 1 component, the investigative project or problem assessment task, done partly at school and partly at home, then again the answer is females. But if the focus is on the high stake Mathematical Methods subject, the subject which often serves as a prerequisite for tertiary courses, and on the

traditional examination formats of CAT 2 and CAT 3 in that subject, then the answer is males. Collectively these data illustrate that the form of assessment employed can influence which group, males or females, will have the higher mean performance score in mathematics. Would the small but consistent differences found in favour of males' mean performance on the NAPLAN papers disappear if the tests were changed from their traditional strictly timed, multiple choice and short answer format to one resembling the CAT 1 requirements?

Changes to the Year 12 assessment procedures in Victoria were introduced in 2000, seemingly in response to concerns about student and teacher workload and to issues related to the authentication of student work for the teacher-assessed CATs. The changes were described by Forgasz and Leder (2001) as follows:

For the three VCE mathematics subjects the assessment changes involve the CAT 1 investigative project task being replaced with (generously) timed, classroom based tasks, to be assessed by teachers but with the scores to be moderated by externally set, timed examination results. It is worth recalling that it was on the now replaced format of CAT 1, the investigative project, that females, on average, consistently outperformed males in all three mathematics studies from 1994 to 1999. Is it too cynical to speculate that this consistent pattern of superior female achievement was a tacit factor contributing to the decision to vary the assessment of the CAT 1 task? It is difficult to predict the longer term effects of the new... assessment procedures on students' overall mathematics performance and study scores. Is there likely to be a return to earlier patterns of superior male performance in mathematics? If so, will this satisfy those who are arguing that males are currently the educationally disadvantaged group? (p. 63)

Indigeneity

That there is no ambiguity about the differences in the performance on the NAPLAN tests between Indigenous and non-Indigenous students is clearly apparent from Tables 1 and 2, and widely emphasized elsewhere. Thomson et al. (2011), for example, examined the 2009 PISA data for Australian students and reported a substantial difference between the average performance of Indigenous and non-Indigenous students on the mathematical literacy assessment component. What message is conveyed by the reporting of these differences?

Gutiérrez (2012) has compellingly used the term “gap gazing” to describe pre-occupation with performance differences between selected groups of students and has argued convincingly that highlighting such differences can be counter-productive and reinforce stereotyping. “In its most simplistic form, this approach points out there is a problem but fails to offer a solution... (T)hat it is the analytic lens itself that is the problem, not just the absence of a proposed solution” (Gutiérrez 2012, p. 31) should not be ignored.

As mentioned earlier, the results of NAPLAN testings are widely disseminated and described in media outlets. Forgasz and Leder (2011) compared the more nuanced reporting of students' results on these tests in scholarly outlets with the

more superficial tone of print media reports. According to these authors “media reports on students’ performance in mathematics testing regimes appear to rely heavily on the executive summaries that accompany the full reports of these data... (T)he more detailed and complex analyses undertaken of entire data sets are often omitted” (p. 218). These comments apply equally to the simplified reporting of gender differences, and differences in performance between Indigenous and non-Indigenous students. It is the arguments advanced in the “more superficial tone of the print media reports” that capture the attention of the general public and shape the sociocultural norms and expectations of the broader society. These norms and expectations are, as mentioned above, among the factors identified by Hyde and Mertz (2009) (among others) as contributing to or averting the emergence of gender difference in performance in mathematics.

Unease has been expressed, both nationally and internationally, about the negative impact of high stake, national testing. Common concerns:

range from the reliability of the tests themselves to their impact on the well-being of children. This impact includes the effect on the nature and quality of the broader learning experiences of children which may result from changes in approaches to learning and teaching, as well as to the structure and nature of the curriculum (Polesel 2012, p. 4).

Disadvantages stemming from blanket reporting of results in large scale examinations have also been widely discussed and selectively elaborated by Berliner (2011). Although his remarks were aimed at indiscriminate and shallow reporting of the PISA results of selected groups of students in the USA, many of his comments are equally applicable to the coverage of performance of Indigenous students on the Australian NAPLAN tests. Three of his concerns seem highly relevant with respect to the portrayal of the numeracy results of Indigenous students: “what was not reported”, “social class”, and “the rest of the curriculum”.

What Was not Reported

Each year the NAPLAN data are published, the rather high proportion of Indigenous students who fail to meet the nationally prescribed minimum numeracy standard attracts the attention of educators and the wider community. As noted by Forgasz and Leder (2011), p. 213:

The lower performance of Indigenous students, compared with the wider Australian school population, attracted sustained media attention. The discovery that Aboriginal students living in metropolitan areas as a group performed almost as well as their non-Indigenous peers received less media attention than the more startling finding that Aboriginal students living in remote communities had an extremely high failure rate of 70–80 %. ‘A combination of low employment and poor social conditions were explanations offered for the distressingly poor performance... their different pass rates are the result of different schooling’ (and a high level of absenteeism).

Aggregating data for all Indigenous students overlooks the large diversity within this group, the range of different needs that inevitably accompany such diversity and

the fact that there are also Indigenous students who perform at the highest level on the NAPLAN test. Pang et al. (2011) identified how valuable data are lost when the performance of a multi ethnic group is described and treated as a single entity, rather than reportedly separately for each constituent group. “Educational policies and statistical practices in which achievement is measured using the (group) aggregate result in over-generalized findings” (p. 384) and hide, rather than identify, the strengths and needs of the different subgroups. These remarks are highly relevant given the many subgroups within the Indigenous community. Gross reporting of achievement outcomes fails to recognize the substantially different backgrounds, locations, needs, and capabilities of individuals within the broader group.

Social Class

There is much diversity in the home background of Indigenous students. Some live in remote areas; others in urbanized centres with access, inside and outside the home, to the same resources as non-Indigenous students. Social class related differences in performance apply to both Indigenous and non-Indigenous students. Although Indigeneity and family background are among the categories reported separately for group results on the NAPLAN test, there is no explicit information about the interactive effects of these variables on performance. To paraphrase Berliner (2011): the scores of Indigenous students, as a group, are likely to remain low, “not because of the quality of its teachers and administrators, necessarily, but because of the distribution of wealth and poverty and the associated social capital that exist in schools” (p. 83) in different metropolitan and remote communities. In the reporting of NAPLAN data for Indigenous students, the emphasis is disproportionately on those performing below expectations without sufficient recognition of confounding, contributing factors, while high performing Indigenous students remain largely invisible.

The Rest of the Curriculum

Under this heading Berliner (2011) focuses particularly on the narrowing of the curriculum, within and beyond mathematics, when the perceived scope and requirements of a national testing program overshadow other considerations and influence the delivery of educational programs. Although this criticism cannot be ignored with respect to the NAPLAN tests, I want to focus here on another, equally pervasive issue.

In recent years, many special programs for Indigenous students have been devised, and implemented with varying degrees of success. Difficulties associated with achieving a satisfactory synchrony between the intended and experienced curriculum for Indigenous students in remote communities have been discussed by Jorgensen and Perso (2012).

In the central desert context, the Indigenous people speak their home languages which are shaped by, and also shape, their worldviews. In Pitjantjatjara, for example, the language is quite restricted in terms of number concepts. The lands of the desert are quite stark with few resources so the need for a complex language for number is limited. As such, the counting system is one of ‘one, two, three, big mob’. It is rare that a collection of three or more occurs so the need for a more developed number system is not apparent. Even when living in community, the need for number is limited. Few people are aware of their birthdates, and numbers in community are very limited in terms of home numbers or prices in the local store. As such, the immersion in number that is common in urban and regional centres is very limited in remote communities. Therefore, many of the taken for granted assumptions about number that are part of a standard curriculum are limited in this context. This makes teaching many mathematical/number concepts quite challenging as it is not only the teaching of mathematical concepts and processes but a process of induction into a new culture and new worldview (Jorgensen and Perso, pp. 127–128).

Many Indigenous students live and learn in conditions more closely aligned to mainstream educational life in Australia than that depicted for Pitjantjatjara. Nevertheless, this snapshot of the prevailing norms and customs of one community highlights factors that will confound a simplistic interpretation of Indigenous group performance data.

NAPLAN and Mathematics Education Research

Not surprisingly, the introduction of NAPLAN has already fuelled a variety of research projects. An overview of work referring substantively to NAPLAN data and presented at the joint conference in 2011 of the Australian Association for Mathematics Teachers (AAMT) and Mathematics Education Research Group of Australasia (MERGA) is summarized in Table 3. It provides a useful indication of the scope and diversity of these investigations.⁵ It is worth noting that the 2011 conference represented the first time the two associations held a fully joint conference. According to Clark et al. (2011) it was a unique opportunity for “practitioners and researchers to discuss key issues and themes in mathematics education, so that all can benefit from the knowledge gained through rigorous research and the wisdom of practice” (p. iii). In addition to “participants from almost every university in Australia and New Zealand, teachers from government and nongovernment schools systems throughout Australia and officers from government Ministries of Education” (Clark et al. 2011), p. iii, there were authors and presenters from a range of other countries.⁶

⁵ Details are extracted from the published proceedings of this joint conference, comprising 130 papers. The proceedings consisted of two sets of papers: *Research papers* and *Professional papers*, reviewed respectively according to established MERGA and AAMT reviewing processes.

⁶ These included Singapore, the United States of America, Papua New Guinea and the United Kingdom.

Table 3 NAPLAN related papers presented at the AAMT-MERGA conference in 2011^a

Author and paper title	Summary of paper and findings/recommendations
Callingham mathematics assessment: everything old is new again?	<p>Descriptive, rather than incisive, reference was made to the NAPLAN testing program in this presentation. Noted were: the contradiction between teachers generally being urged to use formative assessment and the prominence given to the external measure of numeracy provided by NAPLAN; that no significant change has been captured “across time for any grade group” from 2008 to 2010; and that the NAPLAN “results are used for accountability at the local level”. A brief reference is also made to one setting where school based NAPLAN results are used to address elements on which students under-performed</p>
Connolly refining the NAPLAN numeracy construct	<p>An overview is provided of the development of the 2009 and 2010 NAPLAN numeracy test papers. The core content of the test is formally based on the set of nationally agreed curriculum outcomes. Avoided are topics for which there are between state variations in the time of the year they are taught. Items are reviewed multiple times with strong input from key stakeholders. Other factors taken into account in the construction of the test include: item difficulty; cognitive dimension (knowing, applying, and reasoning); item context (abstract or non-abstract); the influence of calculators on content (calculators are not allowed in the Years 3 and 5 papers but at the Year 7 and 9 levels both calculator and non-calculator papers are set); guidelines for item writing; and using accessible language. The Rasch model (Wright, 1980) is used to analyse the test results. This requires not only that certain pre-conditions are met (items are uni-dimensional, locally independent, and uniformly discriminating) but also “allows for sensible comparisons of test scores between different years”</p>
Edmonds-Wathen locating the learner: indigenous language and mathematics education	<p>The author describes the difficulties encountered by Indigenous language speaking students when faced with the typical development of number concepts in the curriculum in the early years of schooling and argues that a different, and group-tailored sequencing of material should be considered. The obstacles created by a “cognitive mismatch between the teacher and student” may fail to gauge accurately the students’ understanding of, for example, spatial items and be reflected in low scores on such items on NAPLAN tests—invaliding simplistic comparisons between Indigenous and non-Indigenous students</p>
Helme and Teese how inclusive is year 12 mathematics?	<p>NAPLAN test data are part of a larger pool of material tapped to explore the mathematics learning experiences and expectations of students at schools in the northern suburbs of Melbourne, but are not discussed per se. Nevertheless the authors’ conclusions are worth noting: “Perceptions of mathematics classrooms and mathematics teachers, and expectations of success, vary according to subject, (student’s) gender and social background”</p>

(continued)

Table 3 (continued)

Author and paper title	Summary of paper and findings/recommendations
Hill Gender differences in NAPLAN mathematics performance	<p>The performance of females and males was compared on items on the Grade 3 and Grade 9 NAPLAN papers for 2008–2010. On each paper, there were some questions on which both groups performed (percentage correct) equally well. When group differences were found they more frequently favoured males than females (e.g., Year 3 paper NAPLAN 2009, no difference on 4 items, females outperformed males on 10 items, males outperformed females on 21 items; Year 9 papers NAPLAN 2010, no difference on 8 items, females outperformed males on 11 items, males outperformed females on 45 items). These trends suggest a “decline in achievement of females as they progress through their schooling”</p>
Hurst connecting with the Australian curriculum: mathematics to integrate learning through the proficiency strands	<p>The scope and demands of NAPLAN tests should not be allowed to dictate the content of the curriculum, nor restrict the instructional strategies used. According to the author, “NAPLAN test scores can greatly assist teachers if they are used appropriately”. Rather than expanding on this theme, the author argues that teachers should “use a constructivist approach to teaching mathematics... (with) an emphasis on rich conceptual understanding as opposed to the mere acquisition of procedural knowledge” and provides some examples that support this theme</p>
Morley Victorian Indigenous Children’s responses to mathematics NAPLAN Items	<p>Using data from the 2008 Years 5, 7, and 9 NAPLAN papers, “whether children of Indigenous background in Victoria, Australia, have different patterns of mathematical responses from the general population” is explored in this paper. Not surprisingly, both groups perform better on high facility than low facility items. Some advantage in favour of Indigenous students is found on the Space strand of the Year 7 paper but less so on the Year 9 paper. At that level, the Algebra strand appeared to be relatively more difficult for Indigenous students</p>
Nisbet national testing of probability in years 3, 5, 7, and 9 in Australia: a critical analysis	<p>The limits of large scale tests are discussed at some length. Often, Nisbet argues, these tests have “a bias towards mechanical processes, and away from problem solving and creativity”. A focus on the probability questions in the 2009 and 2010 NAPLAN numeracy tests for Years 3, 5, 7, and 9 revealed that there were few probability items overall and that only one such item was included in each year level in the 2010 test. Furthermore, Nisbet argued, the scope and aspects of probability probed by the items seemed unacceptably constrained, “with most being multiple-choice items... (and) fundamentally recognition tasks... (to) identify the correct response”</p>

(continued)

Table 3 (continued)

Author and paper title	Summary of paper and findings/recommendations
Pierce and Chick reacting to quantitative data: teachers' perceptions of student achievement reports	The authors use teachers' reactions, to national and school specific information, provided in table and graphical form in a NAPLAN report (usefulness and difficulty of the table and it accompanying annotations and explanations) to gauge the level of teachers' statistical literacy. "Reactions range(d) from those verging on the statistics-phobic... through to deep engagement with the issues". Many teachers preferred the graphical representation, although some welcomed the details provided in the table. Some "reacted strongly about the overwhelming complexity of the data... (and many) expressed uncertainty or confusion over some or all aspects of the data"
Sullivan and Gunningham a strategy for supporting students who have fallen behind in the learning of mathematics	Two items, of different levels of difficulty, from the 2009 NAPLAN Year 9 (no calculator) paper are used to illustrate how poorly some students are performing in mathematics. After this introduction, an out-of-class intervention (the <i>Getting Ready</i> intervention) used to prepare students from Years 3 and 8 for work being taught in their next mathematics lesson, is described. No further reference is made to NAPLAN tests
Tomazos improving mathematical flexibility in primary students: what have we learned?	It is often assumed that schools in higher socio-economic areas with students who perform well on NAPLAN tests do not need to provide extra support for their students. Data from a pilot program at such schools revealed not only that procedural approaches were often used when teaching calculation strategies but also that with "relatively little system input, experienced teachers' classroom practices can be changed" to incorporate greater use of flexible calculation strategies. NAPLAN data were again used as a measure in sample selection, but no further reference is made in the paper to NAPLAN tests
Vale, Davidson, Davies, Hooley, Loton, and Weaven using assessment data: does gender make a difference?	Students' performance on NAPLAN tests was among the measures used to determine a student's learning needs and to select students for specifically designed intervention programs. Gender related differences in performance are reported but no further reference is made to NAPLAN tests
White and Anderson teachers' use of national test data to focus numeracy instruction	The authors argue that, without wishing to advocate 'teaching to the test', much can be gained by teachers who use NAPLAN data from their own school to identify students' numeracy needs and develop instructional strategies to combat faulty practices or inadequate understanding. The approach adopted in one school is described. Whether "professional learning support (had) an impact on student learning and on teaching practice" was also examined

^a To conform with space constraints, the entries in this table are not included separately in the reference list at the end of this paper. All can be found in Clark et al. (2011)

Reference to NAPLAN tests was made in some 10 % of the published papers. As can be seen from Table 3, aspects covered in these papers included issues pertaining to the development of the tests, interpreting the published results of the tests, using test results for curriculum development, and examining the performance of groups of interest, specifically boys and girls and Indigenous students. In some papers reference to NAPLAN data was very much secondary to the core issue explored, for example its (seemingly increasing) use as part of a series of measures to identify a specific group worthy, or in need of, further attention. What could be learnt from the NAPLAN tests about the performance and numeracy needs of high achieving students has, however, not yet attracted research attention. The finding by Pierce and Chick is particularly disturbing. When asked about the statistical and graphical summaries of NAPLAN data relevant to their students the reactions of teachers in their sample ranged “from those verging on the statistics-phobic ... through to deep engagement with the issues”. The NAPLAN national reports contain much valuable and potentially usable data. But how much of these are actually understood and used constructively?

Final Words

After collating information from some 70 public opinion polls in which questions about the efficacy of national tests were included, Phelps (1998) reported:

The majorities in favor of more testing, more high-stakes testing, or higher stakes in testing have been large, often very large, and fairly consistent over the years and across polls and surveys and even across respondent groups (with the exception of some producer groups: principals, local administrators, and, occasionally, teachers) (p. 14) .

The data on which Phelps based his conclusions are now somewhat dated. How the Australian public today values national tests, and in particular the NAPLAN testing regime, is a question still waiting to be investigated. When planning future research activities, whether linked to NAPLAN, to gender and mathematics performance, to issues pertaining to Indigenous students, or to the needs of highly able students, the recommendation of Purdie and Buckley (2010) is well worth heeding:

Although it is important to continue small, contextualised investigations of participation and engagement issues, more large-scale research is called for. Unless this occurs, advancement will be limited because sound policy and generalised practice cannot be extrapolated from findings that are based on small samples drawn from diverse communities (p. 21).

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2010, March). *Australian curriculum information sheet. Mathematics*. Retrieved from http://www.acara.edu.au/verve/_resources/Mathematics.pdf
- ACARA. (2011a). *National assessment program. Literacy and numeracy. Achievement in reading, persuasive writing, language conventions and numeracy*. National report for 2011. Retrieved from http://www.nap.edu.au/_Documents/National%7FReport/NAPLAN_2011%7FNational_Report.pdf
- ACARA. (2011b). *NAP sample assessment*. Retrieved from http://www.nap.edu.au/NAP_Sample_Assessments/index.html
- Berliner, D. C. (2011). The context for interpreting the PISA results in the U.S.A.: Negativism, chauvinism, misunderstanding, and the potential to distort the educational systems of nations. In M. A. Pereyra, H.-G. Korthoff, & R. Cowen (Eds.), *PISA under examination. Changing knowledge, changing tests, and changing schools* (pp. 77-96). Rotterdam, The Netherlands: Sense Publishers.
- Clark J., Kissane, B., Mousley, J., Spencer, T., & Thornton, S. (Eds.) (2011). Traditions and [new] practices. *Proceedings of the AAMT–MERGA conference held in Alice Springs, 3–7 July 2011, incorporating the 23rd biennial conference of The Australian Association of Mathematics Teachers Inc. and the 34th annual conference of the Mathematics Education Research Group of Australasia Inc.* Adelaide, South Australia: AAMT & MERGA
- Cox, P. J., Leder, G. C., & Forgasz, H. J. (2004). The Victorian Certificate of Education – Mathematics, science and gender. *Australian Journal of Education*, 48(1), 27-46.
- Curriculum Corporation. (2006). *Statements of learning for mathematics*. Retrieved from http://www.mceetya.edu.au/verve/_resources/SOL_Maths_Copyright_update2008.pdf
- Dowling, P., & Burke, J. (2012). Shall we do politics or learn some maths today? Representing and interrogating social inequality. In H. Forgasz & F. Rivera (Eds.), *Towards equity in mathematics education. Gender, culture, and diversity* (pp. 87-103). Berlin: Springer-Verlag.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103-127.
- Erchick, D. B. (2012). Preface to “Moving towards a feminist epistemology of mathematics”. In H. Forgasz & F. Rivera (Eds.), *Towards equity in mathematics education. Gender, culture, and diversity* (pp. 9 – 14). Berlin: Springer-Verlag.
- Eurydice (2011). *Mathematics in Europe: Common challenges and national policies*. Brussels: Education, Audiovisual and Culture Executive Agency. Retrieved from http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/132EN.pdf
- Forgasz, H. J., & Leder, G. C. (2001). The Victorian Certificate of Education – a Gendered Affair? *Australian Educational Researcher*, 28(2), 53-66.
- Forgasz, H. J., & Leder, G.C. (2011). Equity and quality of mathematics education: Research and media portrayals. In B. Atweh, M. Graven, W. Secada, & P. Valero (Eds.), *Mapping equity and quality in mathematics education* (pp. 205-222). Dordrecht, the Netherlands: Springer.
- Fyde, J. S., & Meertz, J. E. (2009). Gender, culture, and mathematics. *Proceedings of the National Academy of Sciences of the United States*, 106(22), 8801-8807.
- Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education? In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for mathematics education. Theories, practices, and policies* (pp. 17-33). Dordrecht, The Netherlands: Springer.
- Jacobsen, L. (2012). Preface to “Equity in mathematics education: Unions and intersections of feminist and social justice literature”. In H. Forgasz & F. Rivera (Eds.) *Towards equity in mathematics education. Gender, culture, and diversity* (pp. 31-37). Berlin: Springer-Verlag.
- Joint committee on testing practices. (2004). *Code of fair testing practices in education*. Washington. DC: Joint Committee on Testing Practices. Retrieved from <http://www.apa.org/science/jctpweb.html>
- Jorgensen, R., & Perso, T. (2012). Equity and the Australian curriculum: mathematics. In B. Atweh, M. Goos, R. Jorgensen, & D. Siemon, (Eds.), *Engaging the Australian national*

- curriculum: Mathematics – perspectives from the field* (pp. 115-133). Online Publication: Mathematics Education Research Group of Australasia. Retrieved from <http://www.merga.net.au/sites/default/files/editor/books/1/Chapter%206%7F%20Jorgensen.pdf>
- Kaiser, G., Hoffstall, M., & Orschulik, A.B (2012) Gender role stereotypes in the perception of mathematics: An empirical study with secondary students in Germany. In H. Forgasz & F. Rivera (Eds.), *Towards equity in mathematics education. Gender, culture, and diversity* (pp. 115-140). Berlin: Springer-Verlag.
- Kane, J.,M., & Mertz, J.,E. (2012). Debunking myths about gender and mathematics performance. *Notices of the AMS*, 59(1), 10-20.
- Kenney, P.,A., & Schloemer, C.,G. (2001). Assessment of student achievement, overview. In S. Grinstein & S.I. Lipsey (Eds), *Encyclopedia of mathematics education* (pp. 50-56). New York: RoutledgeFalmer.
- Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 597-622). New York: Macmillan.
- Lundgren, U. P. (2011). PISA as a political instrument. In M. A. Pereyra, H-G. Kotthoff, & R. Cowen (Eds.), *PISA as a political instrument* (pp. 17-30). Rotterdam, The Netherlands: Sense Publishers.
- Northern Territory Government. (n.d.). 2010 NAPLAN analysis using curriculum focus for literacy. Retrieved from http://www.det.nt.gov.au/_data/assets/pdf_file/0005/15827/NAPLAN_Analysis_Curriculum_FocusForNumeracy.pdf
- Pang, V. O., Han, P. P., & Pang, J. M. (2011). Asian American and Pacific Islander students: Equity and the achievement gap. *Educational Researcher*, 40(8), 378-389.
- Phelps, R. P. (1998). The demand for standardized student testing. *Educational Measurement: Issues and Practice*, Fall, 5-23.
- Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The experience of education: The impacts of high stake testing on students and their families*. University of Western Sydney: The Whitlam Institute. Retrieved from http://www.whitlam.org/_data/assets/pdf_file/0008/276191/High_Stakes_Testing_Literature_Review.pdf
- Postlethwaite, T. N., & Kellaghan, T. (2009). *National assessment of educational achievement*. Paris: UNESCO, International Institute for Educational Planning. Retrieved from http://www.iiep.unesco.org/fileadmin/user_upload/Info_Services_Publications/pdf/2009/EdPol9.pdf
- Purdie, N., & Buckley, S. (2010). *School attendance and retention of Australian Indigenous Australian students*. Issues paper no. 1, Closing the gap clearinghouse. Retrieved from <http://www.aihw.gov.au/uploadedFiles/ClosingTheGap/Content/Publications/2010/ctg-ip01.pdf>
- Sidoti, E., & Keating, J. (2012). Foreword. In J. Polesel, N. Dulfer, & M. Turnbull. *The experience of education: The impacts of high stake testing on students and their families* (p. 3). University of Western Sydney: The Whitlam Institute. Retrieved from http://www.whitlam.org/_data/assets/pdf_file/0008/276191/High_Stakes_Testing_Literature_Review.pdf
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, doi: 10.1037/a0026617
- Thomson, S. (2010). *Mathematics learning: What TIMSS and PISA can tell us about what counts for all Australian students*. Paper presented at the ACER Research conference: Teaching mathematics? Make it count. Retrieved from http://research.acer.edu.au/research_conference/RC2010/17august/6/
- Thomson, S., De Bortoli, L., Nicholas, M., Hillman, K., & Buckley, S. (2011). *Challenges for Australian education: Results from PISA 2009*. Retrieved from <http://www.acer.edu.au/documents/PISA-2009-Report.pdf>
- UNESCO (2012). *World atlas of gender equality in education*. Paris: Author.
- Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, doi:10.1016/j.intell.2010.04.006.
- Wright, B. (1980). Foreword. In G. Rasch, *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: The University of Chicago Press.