

# NAPLAN Review Final Report

August 2020

Barry McGaw  
William Loudon  
Claire Wyatt-Smith



## Copyright

© State of New South Wales (Department of Education), State of Queensland (Department of Education), State of Victoria (Department of Education and Training), and Australian Capital Territory, 2020.

Subject to the exceptions listed below, the material available in this publication is owned by the State of NSW, State of Queensland, State of Victoria and Australian Capital Territory and is protected by Crown Copyright in each state/territory. It is licensed under the [Creative Commons Attribution 4.0 International Licence](#). The legal code for the licence is available [here](#).



## Attribution

### NAPLAN Review Final Report

Emeritus Professor Barry McGaw AO, Emeritus Professor William Loudon AM and Professor Claire Wyatt-Smith.

© State of New South Wales (Department of Education), State of Queensland (Department of Education), State of Victoria (Department of Education and Training), and Australian Capital Territory, 2020.

## Exceptions

The Creative Commons licence does not apply to:

1. The logos of any of the copyright owners or their departments
2. The Coat of Arms of Australia or a State or Territory of Australia
3. Material owned by third parties that has been reproduced with permission. Permission will need to be obtained from third parties to re-use their material.

If you have questions about the copyright in the content of this website, please contact the NSW Department of Education on 1300 679 332 or email [DoEinfo@det.nsw.edu.au](mailto:DoEinfo@det.nsw.edu.au).

## ISBN

978-0-6480638-1-0

## Acknowledgment of Country

We acknowledge the homelands of all Aboriginal people and pay our respect to Country.

**The Hon. Sarah Mitchell**  
Minister for Education &  
Early Childhood Learning  
(NSW)

**The Hon. James Merlino**  
Minister for Education  
(Vic)

**The Hon. Grace Grace**  
Minister for Education  
(Qld)

**Ms Yvette Berry**  
Minister for Education  
& Early Childhood  
Development (ACT)

Dear Ministers

In September 2019, you commissioned us to review the National Assessment Program: Literacy and Numeracy (NAPLAN). It was an honour and a pleasure to undertake this important work.

NAPLAN has been in place since 2008 and in this time, as you well know, it has received mixed reactions from stakeholders. You asked us to take account of the changing local and international education landscapes and to consider the extent to which NAPLAN remains fit-for-purpose.

As you directed in our terms of reference, we have identified what a standardised assessment regime in Australian schools should deliver, determined how well NAPLAN achieves this, and identified short- and long-term improvements that can be made.

We consulted widely, despite interruptions from the COVID-19 pandemic, meeting with 175 stakeholders in 91 meetings and receiving 301 responses to an online survey and an additional 31 written submissions. We formed a Practitioners' Reference Group made up of teachers, principals and a union representative for more extended discussion on a number of important areas. We also collaborated with numerous international colleagues to investigate practices in several high-performing countries. We are grateful for the time each of these contributors so generously gave and we have acknowledged them in our report.

We have concluded that standardised assessment is important in Australian education and that it serves a variety of purposes. We have recommended the retention of some important features of NAPLAN but recommended, as well, some important changes to content, psychometric properties, timing of the assessments within the year and over the Years of schooling.

We commend our final report to you and trust that it will provide a useful platform for a revitalised Australian National Standardised Assessment.

Yours sincerely



**Emeritus Professor  
Barry McGaw AO  
Chair**



**Emeritus Professor  
William Loudon AM**



**Professor  
Claire Wyatt-Smith**

14 August 2020

## Contents

Copyright .....	2
Attribution.....	2
Exceptions.....	2
ISBN.....	2
Acknowledgment of Country .....	2
<b>Executive Summary .....</b>	<b>8</b>
Introduction .....	8
Nature of standardised assessments.....	8
Standardised assessment in Australia .....	8
Concerns about publication of school results .....	9
No common assessment practices in high-performing countries.....	9
Linking standardised assessments to the curriculum .....	10
Moving NAPLAN Online with branching tests.....	10
Problems with NAPLAN writing test.....	11
Participation rates in NAPLAN .....	11
Census or sample assessment .....	11
Recommendations for change.....	12
<b>Preface.....</b>	<b>14</b>
Context.....	14
Task and timeframe of the review .....	14
Summary of key dates .....	15
Process.....	15
Acknowledgements.....	16
<b>Chapter 1: Purposes of standardised assessment.....</b>	<b>17</b>
Standardised assessment.....	18
Large-scale standardised assessment in Australian schools.....	18
Purposes of national standardised testing .....	19
Summary.....	23
Current purposes of the national standardised testing program.....	23
Stakeholder perspectives on the purposes of national standardised assessment.....	24
<b>Chapter 2: National and international measures of achievement .....</b>	<b>28</b>
NAPLAN achievement and improvement, 2008 to 2019 .....	29
Reading.....	29
Writing.....	30
Spelling .....	32
Grammar and punctuation.....	33
Numeracy.....	34
Patterns of change across the NAPLAN test domains.....	35
NAPLAN and international surveys of student achievement.....	36
PIRLS and NAPLAN reading .....	37
TIMSS and NAPLAN numeracy.....	38
PISA reading literacy and NAPLAN reading .....	39

PISA mathematical literacy and NAPLAN numeracy .....	39
The national and international standardised testing programs compared .....	40
Stakeholder views on national and international standardised testing .....	42
<b>Chapter 3: Other national educational assessment practices .....</b>	<b>44</b>
Country assessment policies and practices .....	45
Singapore .....	45
Japan .....	46
Canada – Ontario .....	47
England .....	48
Scotland .....	49
New Zealand .....	51
Finland .....	52
Potential relevance for Australia .....	53
<b>Chapter 4: Quality of NAPLAN digital tests .....</b>	<b>57</b>
Content of tests .....	58
Paper tests .....	58
Online tests .....	60
Item selection .....	63
Links to the Australian Curriculum .....	65
Psychometric properties of the tests .....	66
Effect of branching within online tests .....	66
Scaling of results over year levels and time .....	68
Confidence in measurement .....	70
Establishing benchmarks .....	73
Inclusiveness of the tests .....	75
Students with disability .....	77
Aboriginal and Torres Strait Islander students .....	78
Cultural and language diversity .....	79
Similarity to other tests used in schools .....	79
Summary .....	81
<b>Chapter 5: Quality of the NAPLAN writing test .....</b>	<b>83</b>
NAPLAN writing test .....	84
Critique of the NAPLAN writing test .....	86
Formulaic teaching of writing and teaching writing as formulaic .....	87
Factors internal and external to the test .....	88
The writing test and alignment to the Australian Curriculum .....	89
Sex .....	90
Geographic location .....	93
High performance in writing at Year 9 .....	95
Critique of the writing test .....	96
NAPLAN writing test mode .....	100
Summary .....	101

<b>Chapter 6. Uses of NAPLAN.....</b>	<b>103</b>
National uses.....	104
School systems/sectors.....	107
Schools.....	109
Individual teachers.....	109
Family and community.....	110
Summary.....	112
<b>Chapter 7: Recommendations.....</b>	<b>113</b>
National standardised assessment.....	113
Purposes of national standardised assessment.....	113
Features of an assessment system.....	114
Role of NAPLAN in meeting national purposes.....	118
Changes to the NAPLAN tests.....	120
Curriculum coverage.....	120
Frequency and timing of tests.....	123
Rebranding the program.....	125
Redeveloping the online branching tests.....	126
Redeveloping the writing test.....	128
Starting a new time series.....	132
Reporting.....	132
Monitoring trends.....	132
Reporting to schools, parents/carers and the community.....	133
Ongoing evaluation.....	134
Links to terms of reference and proposed timeline.....	135
Appendix 1: Summary of recommendations.....	142
Appendix 2: Review of NAPLAN terms of reference.....	146
Background.....	146
Terms of reference.....	146
Other relevant work that the review will need to consider.....	147
Review process.....	147
Review outputs.....	147
Appendix 3: List of stakeholders consulted.....	148
Stakeholder consultations.....	148
Appendix 4: International practice in standardised writing assessment.....	153
International testing of writing.....	153
Summary.....	162
<b>References.....</b>	<b>168</b>

## List of tables

<b>Table 1:</b> Census and sample assessment and the purposes of national standardised assessment.....	12
<b>Table 2:</b> Differences in achievements of students in reading, 2008 to 2019 .....	29
<b>Table 3:</b> Differences in achievements of students in writing, 2011 to 2019.....	31
<b>Table 4:</b> Differences in achievements of students in spelling, 2008 to 2019 .....	32
<b>Table 5:</b> Differences in achievements of students in grammar and punctuation, 2008 to 2019.....	33
<b>Table 6:</b> Differences in achievements of students in numeracy, 2008 to 2019.....	34
<b>Table 7:</b> Differences in achievement, base year to 2019, by test domain, Australia.....	35
<b>Table 8:</b> Differences in achievement, base year to 2019, by test domain, Western Australia .....	36
<b>Table 9:</b> Differences in achievement, NAPLAN, PIRLS, TIMSS and PISA .....	41
<b>Table 10:</b> Position of comparison countries in relation to Australia in PISA 2018 .....	54
<b>Table 11:</b> Nature of assessments in other countries.....	55
<b>Table 12:</b> Structure of paper NAPLAN paper tests, 2019.....	59
<b>Table 13:</b> Structure of NAPLAN Online tests, 2019.....	64
<b>Table 14:</b> 2019 scales for which distributions were adjusted to match those for 2017 .....	70
<b>Table 15:</b> Confidence ranges (95%) for the 2019 NAPLAN Year 5 numeracy scores.....	71
<b>Table 16:</b> Percentages of Australian students below minimum standards benchmarks.....	75
<b>Table 17:</b> Percentages of non-participating students in NAPLAN 2017 tests .....	76
<b>Table 18:</b> Participation rate (%) in NAPLAN in 2017 .....	76
<b>Table 19:</b> Percentage of male and female students below National Minimum Standard in writing .....	91
<b>Table 20:</b> Percentage of students by location below National Minimum Standard in writing .....	92
<b>Table 21:</b> Descriptions of performance bands on the writing scale .....	94
<b>Table 22:</b> Percentages of Year 9 students in top two bands in NAPLAN writing.....	96
<b>Table 23:</b> Census and sample assessment and the purposes of national standardised assessment .....	114
<b>Table 24:</b> Terms of Reference, recommendations and timeline .....	135
<b>Table 25:</b> Model for assessment of writing in multiple languages.....	162
<b>Table 26:</b> International large-scale assessment of writing.....	163

## List of figures

<b>Figure 1:</b> NAPLAN Review summary of key dates.....	15
<b>Figure 2:</b> Branching structure of NAPLAN Online literacy and numeracy tests .....	60
<b>Figure 3:</b> Branching structure of NAPLAN Online grammar and punctuation test.....	61
<b>Figure 4:</b> Proportions of students taking each path – Year 3 numeracy, 2019 .....	66
<b>Figure 5:</b> Distributions of student achievement by pathway – Year 3 numeracy, 2019.....	67
<b>Figure 6:</b> Extent of uncertainty in student NAPLAN results with print and branching digital tests, 2018 .....	72
<b>Figure 7:</b> NAPLAN assessment scale .....	74
<b>Figure 8:</b> My School comparison with students with same starting score and similar background .....	105
<b>Figure 9:</b> My School: Selected school compared with students with and similar background.....	105
<b>Figure 10:</b> Relationship between schools' socio-educational advantage and NAPLAN results .....	107
<b>Figure 11:</b> Trends in mean performances on the NAPLAN Writing test.....	129
<b>Figure 12:</b> Categories of validity evidence in large-scale assessment of writing .....	154

# Executive Summary

## Introduction

The NAPLAN Review has been commissioned to determine what the objectives of national standardised assessment should be, to advise on how well the National Assessment Program: Literacy and Numeracy (NAPLAN) meets these objectives and how NAPLAN compares with national assessment programs in other countries, and to identify short- and longer-term improvements in the national standardised assessment of literacy and numeracy. NAPLAN was built on almost two decades of similar testing at a state and territory level and was introduced in 2008. NAPLAN has evolved during the period since its introduction but it is timely to take stock and identify whether and how it might now be changed.

This report begins with a discussion of the history and purposes of standardised testing in Australia. It then considers trends in achievement revealed by NAPLAN and international studies in which Australia participates (Chapter 2) and other national assessment policies and practices (Chapter 3). The strengths and weaknesses of the current NAPLAN assessments are examined in reading, language conventions and numeracy (Chapter 4) and writing (Chapter 5). Chapter 6 discusses the uses made of NAPLAN by systems/sectors, schools, teachers and parents. Chapter 7 proposes a series of short- and longer-term improvements to Australia's national standardised assessment program.

## Nature of standardised assessments

Standardised assessments provide common test-taking conditions, questions,

time to respond, scoring procedures and interpretations of the results. They may test knowledge, skills, attributes or values. They may use multiple choice, short answer or extended written responses. They may be administered to whole populations, to samples from a population or to individuals. They may be marked electronically or by human assessors. Results may be reported in terms of standards achieved or in comparison with the achievements of a wider population. Results may be used for summative, formative, diagnostic or predictive purposes. There is a long history of standardised testing in Australia, beginning with junior and senior secondary school examinations conducted by universities and state-wide examinations at the end of primary school. There is also a range of commercially available standardised tests that schools and teachers use to measure students' achievements and monitor their progress.

## Standardised assessment in Australia

The abolition of external examinations before the end of secondary education meant that parents had no standardised assessments and reports on children's progress from a broader perspective than that of their children's own school. Schools could still use standardised tests and their published norms but that did not provide parents with the kind of comparisons across the age group that the external examinations provided. In that vacuum, all states and territories over a period from the late 1980s to the mid-1990s introduced new tests that were administered to all children in several year levels in their movement through school. These census tests were limited to literacy and numeracy though



some states and territories also conducted sample surveys of students' achievement and progress in other curriculum areas. Once all states and territories were assessing all students in literacy and numeracy, the ministerial council sought a national perspective from the different jurisdictions' results. In 2007, they resolved to use common tests, selected the name National Assessment Program: Literacy and Numeracy (NAPLAN) and introduced the new tests in 2008.

Since then, five purposes for national standardised testing have been endorsed by the ministerial council:

- Monitoring national, state and territory programs and policies.
- System accountability and performance.
- School improvement.
- Individual student learning achievement and growth.
- Information for parents on school and student performance.

NAPLAN has been a useful barometer with which to examine trends in students' achievements over time. In the period 2008 to 2019, national NAPLAN results have revealed improvement in reading and numeracy in primary schools but not in secondary schools, static performance in writing in Years 3 and 5 and a decline in Years 7 and 9. They have also revealed different patterns among the states and territories. Queensland and Western Australia have improved more than the others, but they started behind the ACT, NSW and Victoria and have not surpassed them. At the same time as testing all students through NAPLAN, Australia has participated in a number of international sample surveys of students' achievements in literacy and numeracy all of which, while showing improvement in some year levels in some domains, have generally

produced declining results and shown a larger proportion of Australian students than in NAPLAN to be below the levels that each assessment program sets as defining minimum competence. The international surveys have also reported a declining proportion of high performing students.

## Concerns about publication of school results

There was some resistance among teacher organisations to the initial introduction of state and territory census assessments of students in literacy and numeracy but that had largely dissipated until, in 2010, the My School website was introduced and provided public reporting on schools' NAPLAN results. While the website provided only comparisons among schools with students from similar levels of socio-educational advantage, many newspapers retrieved data to create raw league tables that ranked schools without consideration of differences in context. It also meant that schools were being compared and publicly ranked on only the narrow criteria of students' literacy and numeracy. That has led many in public debate and in submissions to this and earlier reviews to campaign for NAPLAN to assess only samples of students, sufficient to monitor national, state and territory and other trends in students' achievement levels but not to enable inter-school comparisons to be made.

## No common assessment practices in high-performing countries

A review of practices in high-performing countries revealed no common patterns in assessment policies and practices. In Finland and New Zealand there are only sample assessments to monitor the education system, with teachers' professional judgements being the basis of reporting to parents and students. In Scotland,

new census assessments were introduced in 2017 to 2018, though schools have the right to opt out. The student participation rates were 95% in 2017 to 2018 and 93.4% in 2018 to 2019. These participation rates match those achieved with NAPLAN census tests in Australia. Results are reported to schools, with student reports similar to those provided to Australian parents from NAPLAN, but in Scotland they are intended to be used only by schools and teachers as one piece of evidence contributing to reports to parents/carers, students and local education authorities.

Ontario, the most populous Canadian province, conducts census assessments in literacy and numeracy with reports to parents and public reporting of schools' results. In England, there are census assessments of literacy and numeracy in primary and lower secondary education. Reporting is focused, as it is on My School in Australia since 2019, on growth not current achievement levels. In mid-secondary school (Years 10 to 11, aged 16), England has retained subject-based external examinations through its General Certificate of Education Ordinary-Level. Japan has census assessments in Grade 6 and Grade 9 in Japanese, mathematics and science, with English added in 2019. At the end of lower secondary education (Grade 9) there are extremely competitive entrance examinations for senior high schools.

In Singapore, there are no census assessments in general domains like literacy and numeracy. There are subject-based examinations: the Primary School Leaving Examination at the end of primary school, which also serves as a selection test for secondary school, and the Singapore-Cambridge General Certificate of Education examinations in Year 11 prior to end-of-secondary examinations in Year 13.

## Linking standardised assessments to the curriculum

When NAPLAN tests were first developed, there was no common curriculum in Australia, so they were based on national Statements of Learning for English and national Statements of Learning for Mathematics. From 2017, the NAPLAN tests have been based on the Australian Curriculum, but on the literacy and numeracy continua since both are expected to be developed across the subject curricula and not exclusively in English and mathematics. That makes less difference in numeracy/mathematics where there is more overlap than in literacy/English but, in both cases, it is not clear where ownership of both literacy and numeracy lies in a secondary school where teaching and learning are subject-based.

## Moving NAPLAN Online with branching tests

Since 2018, schools have increasingly taken the NAPLAN tests online. These online tests in numeracy and literacy are adaptive. Students are marked by the computer as they answer questions and branched one third and two thirds of the way through the tests to more or less complicated questions. This branching is determined based on their achievement on the test to the point of branching. That presents students with questions better targeted to their achievement levels. It has resulted in better assessment over the full range of achievement levels among students and less uncertainty in the measurement throughout the range. The level of uncertainty, as with all educational measurement, depends on how much data lies behind a result: national means have the least uncertainty and individual student's results the most. The means for large schools have less uncertainty than those for small schools.

## Problems with NAPLAN writing test

The NAPLAN writing test is the most problematic. The restriction of the writing genres to narrative and persuasive, with the specific genre being announced in advance in the early years of NAPLAN, has led to very formulaic writing in students' responses to the prompt and, as a further unintended consequence, to very formulaic teaching of writing in some schools as they seek to prepare students for the NAPLAN writing test. The marking criteria also need to be reviewed. The language conventions test in spelling has a reliability similar to those achieved in the reading and numeracy tests but the grammar and punctuation test has a markedly lower reliability. A more serious consideration is whether it is better to assess these language conventions with decontextualised test questions or to assess them through students' use of them in their writing. To achieve all of these changes, it is recommended that the writing test be withdrawn from census testing and conducted as a sample survey during a period of experimental redevelopment.

As the NAPLAN tests are moved online, particularly the writing test with its extended response, it will be essential that students develop fluency in the use of keyboards and word processors, at least from Year 5, to enable them to concentrate on the substance of their writing. For Year 3, students' handwritten responses would be more appropriate.

## Participation rates in NAPLAN

There are explicit provisions for some students not to participate in NAPLAN testing. Students can be excluded if their language background is other than English and they have been in Australia for less than a year or they have significant disabilities. In their submissions to the review, parents of children with learning difficulties complained that some schools encouraged their children to be withdrawn. Their preference was to obtain the external measure on achievement with which NAPLAN would enable them to see where their children stood in their age-group. Parents can apply to have their children withdrawn on the basis of religious beliefs or philosophical objections to testing. Beyond those reasons, there is generally a greater, though still small, percentage of the age group that is simply absent on the day of testing. It would be good for jurisdictions to investigate students' reasons for absence and seek to reduce the current levels.

## Census or sample assessment

Many submissions to the review declared that NAPLAN had been introduced for the specific, narrow purpose of monitoring the overall education system but all five of the purposes of standardised assessment nominated above can be found in declarations of the ministerial council about the purposes and uses of NAPLAN results. These purposes are also reflected in current practices of government education departments and some, though not all, schools. Table 1 shows what can be achieved with census and sample testing.

**Table 1:** Census and sample assessment and the purposes of national standardised assessment

Purpose of national standardised assessment	Census	Sample
Monitoring progress towards national goals		
• National, jurisdictional and system estimates of achievement	✓	✓
• Relative performance by gender, geographic location of schools, socioeconomic background and Aboriginal and Torres Strait Islander background	✓	✓
School system accountability and performance		
• Accountability for system performance	✓	✓
• Accountability for school performance	✓	✗
School improvement		
• School-level information on achievement and growth by assessment domain	✓	✗
• School-level targets informed by system comparative data	✓	✗
Individual student learning achievement and growth		
• Student level achievement estimates for comparative purposes (cohort, test domain, gain, equity groups)	✓	✗
• Student level achievement estimates for diagnostic purposes	✗	✗
Information for parents on school and student performance		
• Individual student achievement	✓	✗
• Relative school performance	✓	✗

This table makes clear which current practices could not be sustained if sample rather than census assessment were used.

## Recommendations for change

The development of the full set of recommendations is described in Chapter 7 and they are listed in Appendix 1. Only their main features are included in this Executive Summary.

It is recommended that NAPLAN remain as a census test of students' achievement but that they be taken by students in Years 3, 5, 7 and 10 not Years 3, 5, 7 and 9. Students' achievement levels and the absenteeism rate at Year 9 reveal a relatively low level of student engagement with NAPLAN

compared with Years 3, 5 and 7. In Year 10 students are more mature and, more importantly, reaching the stage at which important choices are to be made about their studies in the upper secondary years. Having a NAPLAN assessment in Year 10 would provide some data to inform discussions between the students and their teachers.

It is also recommended that the tests be administered as early as feasible in the school year, and not in May as at present, to give NAPLAN a more clearly formative role as a measure of students' starting points for the year.

It is recommended that the writing test be redeveloped with richer prompts, removal of restrictions on the genre in which students write, assessment of language conventions of spelling, grammar and punctuation in the students' writing not in stand-alone tests, and inclusion of teacher judgements as a component in the marking. To achieve this substantial change, it is recommended that the writing test be withdrawn from the census assessment while experimental redevelopment is undertaken and replace it with a national sample of schools and students. Once the new form has been established, it is recommended that census assessment of writing be resumed.

It is recommended that the scope of the national standardised assessment be broadened beyond its current limitation to literacy and numeracy. As noted in Chapter 3, a number of countries include science in their census assessments and, as noted in Chapter 2, Australia's performance in science in the international surveys has declined in recent years. In Australia, there is a strong focus in current discussions of curriculum on STEM (science, technology, engineering and mathematics). In the Australian Curriculum, this is best represented by the subjects: science, digital technologies and mathematics. There is also attention being given to the general capabilities in the Australian Curriculum. These are not generic capabilities devoid of subject content. Indeed, they take different forms in different domains. Critical and creative thinking in science, for example, is not the same as critical and creative thinking in history. It is recommended, therefore, that a new test of critical and creative thinking in STEM be introduced at Years 5, 7 and 10. It is recommended that the current triennial sample survey of science literacy in Years 6 and 10 be withdrawn and that consideration be given to its replacement

in the triennial cycle by another that covers both a subject and a general capability from the Australian Curriculum, such as history and intercultural understanding.

Once NAPLAN Online is fully implemented, with the digital tests freed from the restriction of mirroring print forms and capitalising on the flexibility and creativity available in the digital form, it is recommended that a new time series be commenced without reference back to the 2008 NAPLAN scales. For all but the writing test, the full move to online delivery should expedite the return of results to school within days of the testing. Results of the writing test would take longer because of the need for manual marking.

It is recommended that Ministers emphasise the significance of changes that they introduced to My School in 2019 to remove inter-school comparisons of the levels of students' achievements and to focus on students' growth in comparison with other students at the same starting point.

At present the National Assessment Program (NAP) is an umbrella title for both NAPLAN's census tests of students' literacy and numeracy and sample surveys in some other domains. To distinguish them more clearly and to recognise that the census tests are proposed to move beyond literacy and numeracy, it is recommended that programs be rebranded, with new names adopted for each program: the Australian National Standardised Assessments (ANSA) instead of NAPLAN and National Sample Assessment Program (NSAP) instead of NAP.

# Preface

## Context

The National Assessment Program – Literacy and Numeracy (NAPLAN) is an annual, point-in-time assessment undertaken by Australian students in Years 3, 5, 7 and 9. The first NAPLAN tests were conducted in 2008.

The Australian Curriculum, Assessment and Reporting Authority (ACARA) is responsible for the development and central management of the tests. Test administration authorities in each state and territory are responsible for the administration of the tests in their jurisdiction. All states and territories administer the tests in compliance with nationally agreed protocols and are also responsible for marking the tests in accordance with strict guidelines and processes.

NAPLAN tests four areas ('domains') — reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. The tests are scheduled in the second full week of May. In the past, the tests have been conducted in pen and paper format. Schools are currently transitioning to online testing, with more than 50% of schools across Australia participating in NAPLAN Online in 2019 (ACARA, 2020k). Full transition of schools to NAPLAN Online is expected by 2022.

Results are provided to schools between August and September. Every student who participates in the tests receives an Individual Student Report of their results. Since 2010, NAPLAN performance of schools in Australia is reported on the My School website. This includes comparative performance to similar students and national results. In addition, the website displays schools' historical performance and also provides some demographic and financial information for each school.

## Task and timeframe of the review

On 12 September 2019, the NSW Minister for Education and Early Childhood Learning announced that the NSW, Victorian, Queensland and ACT governments would sponsor a review of NAPLAN. Emeritus Professor Barry McGaw, Emeritus Professor William Loudon and Professor Claire Wyatt-Smith were appointed to the NAPLAN Review's panel as independent curriculum and assessment experts.

The panel was asked to identify what a standardised testing regime in Australian schools should deliver, assess how well NAPLAN achieves this, and identify short- and longer-term improvements.

The full terms of reference are set out in Appendix 2.

## Summary of key dates



**Figure 1:** NAPLAN Review summary of key dates

## Process

The review was conducted in two stages. As part of the stage one consultation process, panel members met with 56 individuals over 32 meetings. Meetings were held from Tuesday 22 October to Friday 25 October 2019 in Brisbane, Melbourne, Sydney and Canberra. The panel also drew from other recently completed work, including the 2018 Queensland NAPLAN Review and the 2019 NAPLAN Reporting Review. An interim report setting out some of the major concerns expressed by stakeholders and some preliminary discussion on strategies to deal with these issues was publicly released on 6 December 2019.

In the second stage of the review, the panel further analysed the issues discussed in the interim report and broadened its consultation to hear directly from a greater proportion of stakeholders and experts. This final report discusses the themes raised in consultations and identifies the challenges NAPLAN has faced and potential opportunities moving forward. It also offers a strategic blueprint for the future of standardised assessment in Australia.

Stage two stakeholder consultations were held from 24 March to 27 May 2020. Face-to-face meetings were scheduled in each state but due to the COVID-19 pandemic, these were shifted to a web conference platform. The panel members met with 160 individuals across 53 meetings.

Stakeholders consulted in stages one and two of the review included teachers, principals, parents and carers, students, school systems/sectors, unions, accreditation authorities, teacher subject associations, teacher professional associations, principals' associations, Aboriginal and Torres Strait Islander representative groups, disability and inclusion representative groups, ministers, key assessment bodies, educational organisations, academics, and national and international educational experts.

A full list of stakeholders is available at Appendix 3. Stakeholders consulted as part of both stages one and two of the review have only been listed once.

A call for public submissions for stage two of the review was made in late February 2020. Stakeholders were invited to complete an online survey via targeted invitation as well as via newspaper, web and social media

channels, to ensure all interested parties had an opportunity to contribute. The survey consisted of seven topical questions, as well as some demographic questions. There were 301 responses to the survey from all jurisdictions except Tasmania and the Northern Territory. In addition, the panel received 31 written submissions made by individuals or organisations via email.

A Practitioners' Reference Group was established with 17 teacher/principal representatives from all systems/sectors in each of the four participating states and territories. The panel conducted six meetings with the reference group (two whole group meetings and one meeting with members in each jurisdiction). The group provided the panel with an in-depth practitioner perspective on a range of issues, including test administration, timing and student engagement, data use and classroom practice.

Contributions from the Practitioners' Reference Group, responses to the online survey, written submissions and web conference meetings assisted the panel significantly in their consideration of the challenges and opportunities presented by NAPLAN. Quotes and exploration of the review's consultation have been referenced as part of the panel's analysis of issues throughout this final report.

Dr Sue Thomson from the Australian Council for Educational Research was commissioned to prepare a comparative report on the key differences between what is assessed in international assessments – Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) and any apparent consequential differences in outcomes. The panel have used this information in their final report.

## Acknowledgements

We wish to thank all those who completed our online survey, prepared submissions for our consideration, and those who met with us face-to-face before the COVID-19 restrictions and in web-based meetings after them.

We are particularly grateful to those who provided statistical advice and, in some cases, undertook statistical analyses for us: Dr Raymond Adams, Dr Eveline Gebhardt, Dr Goran Lazendić and Dr Lucy Lu.

We thank colleagues in the Queensland Department of Education who prepared background information on the assessment systems in other countries and the following colleagues who provided advice and reviewed our draft materials – Charles Darr, Associate Professor Christopher DeLuca, Dr Jenny Donovan, Professor Nikolaj Elf, Professor Karen R. Harris, Professor Louise Hayward, Christine Jackson, Professor Kim Koh, Shumpei Komura, Associate Professor Ricky Lam, Associate Professor Yuen Yi Lo, Dr Tim Oates, Professor Judy M. Parr, Dr Poon Chew Leng, Professor Pasi Sahlberg, Professor Gustaf B. Skar, Dr Sue Thomson and Peter Titmanis. Any deficiencies in our descriptions are ours but their help diminished the risk.

We are enormously grateful to Claire Todd, Susanna Osborne and Carolyn Burns in the NSW Department of Education who provided the highly professional Secretariat that supported and, in many skilled and subtle ways, guided us throughout and did everything on time and to the highest quality. We extend that gratitude to the Secretariat members from Victoria, Queensland and the ACT.



# Chapter 1: Purposes of standardised assessment

This chapter begins with a summary of standardised testing in Australia and traces the purposes of the NAPLAN national standardised testing program through the deliberations of successive national education ministerial councils and their Hobart, Adelaide, Melbourne and Alice Springs (Mparntwe) declarations. It concludes with a discussion of these purposes in the context of feedback from stakeholders consulted in this review.

## Key points

- There is a long history of standardised testing in Australia, beginning with junior and senior secondary school examinations conducted by universities.
- State- and territory-based standardised literacy and numeracy tests were developed in the 1980s and 1990s. Subsequently, efforts were made to establish comparability of these tests using benchmarks and statistical equating.
- In 2003, national sample-based testing of science literacy, civics and citizenship, and information and communications technology commenced for Years 6 and 10, with one area tested each year on a triennial cycle.
- National, whole-cohort assessment of literacy and numeracy in Years 3, 5, 7 and 9 began in 2008.
- Schools use a wide variety of other opt-in or compulsory standardised assessments, including early years assessments in every state and territory.
- Through successive ministerial council decisions and declarations, NAPLAN testing has developed five endorsed purposes: monitoring progress towards national goals, school system accountability and performance, school improvement, individual student learning achievement and growth, and information for parents/carers on school and student performance.
- Notwithstanding these officially endorsed purposes, there remain substantial concerns among stakeholders about NAPLAN's capacity to meet all five purposes equally well and without conflict among the purposes.

## Standardised assessment

The essential characteristic of standardised testing is consistency. Standardised tests provide common test-taking conditions, questions, time to respond, scoring procedures and interpretations of the results. Contemporary standardised tests are developed with careful attention to fairness, reliability and validity, and may vary test conditions in order to eliminate bias against individuals or groups of candidates.

Standardised tests take many forms. They may test knowledge, skills, attributes or values. They may use multiple choice, short answer or extended written responses. They may be administered to whole populations, to samples from a population or to individuals. They may be marked electronically or by human assessors. Results may be reported in terms of standards achieved or in comparison with the achievements of a wider population. Such results may be used for summative, formative, diagnostic or predictive purposes.

Australia's national standardised literacy and numeracy tests – NAPLAN – are just one of the possibilities in this universe of standardised tests. They are whole-cohort tests, they focus on cognitive skills in literacy and numeracy, are marked by a combination of digital and expert assessors, report against national standards and are used for a variety of summative and formative purposes.

### Large-scale standardised assessment in Australian schools

The first standardised tests to be used in Australian school education were state-based public examinations, which have been held for more than 150 years. The University of Melbourne established a matriculation examination in 1855, and the University of Sydney set and marked its first junior and senior secondary public examinations in 1867. Responsibility for examinations

such as these subsequently moved from the universities to public examinations boards. Public examinations at the junior secondary level were discontinued but most jurisdictions have continued to use large-scale standardised examinations as a component of their senior secondary certification.

In the absence of public examinations until the end of secondary schooling, during the 1980s a number of jurisdictions introduced large-scale standardised testing in the primary or junior secondary years. The Victorian Achievement Studies were conducted in 1988 and 1990. In 1989, NSW introduced the Basic Skills Tests of literacy and numeracy in Years 3 and 6. In 1990, Queensland conducted the Assessment of Student Performance in Years 5, 7 and 9; and Western Australia began the sample-based Monitoring Standards in Education Program in Years 3, 7 and 10. Although these tests assessed similar skills in literacy and numeracy, there were no common standards across the assessments. In 1997, national benchmarks were established using state-based assessments and a procedure was developed to equate state and territory assessments and seek comparable reporting of results. National assessment and reporting using equated state and territory standardised tests in Years 3 and 5 literacy began in 1998, followed by numeracy testing in 1999.

National sample assessments began with science literacy in 2003, civics and citizenship in 2004 and information and communications technology literacy in 2005. Trials of the first national literacy and numeracy assessments using a common test were announced in 2005. Following evaluation of this trial, whole-cohort national literacy and numeracy using common tests in Years 3, 5, 7 and 9 began in 2008. Under the banner of NAPLAN, these census tests have continued since then.

Beyond NAPLAN, standardised tests are used widely in Australian schools. Some standardised tests are mandated by school systems/sectors, and others are selected and used on an opt-in basis by schools for individuals or groups of students. Most public school jurisdictions, for example, use some form of common monitoring instrument or standardised assessment in the early years of schooling. Students in all NSW public schools are assessed in the Foundation year using the Best Start Kindergarten Assessment. Victorian public schools assess all Foundation students using English Online and Mathematics Online interviews. Queensland public schools have access to the Early Start assessment of literacy and numeracy skills in Years Foundation to 2. ACT public schools use the University of Western Australia's BASE (formerly PIPS) assessments at the beginning and end of the Foundation year. Other schools and jurisdictions use a variety of assessments on an opt-in basis in the junior primary years, including most commonly the Australian Council for Educational Research (ACER) Progressive Achievement Tests in reading and mathematics, BASE and the South Australian Spelling Test.

## **Purposes of national standardised testing**

NAPLAN tests are technically similar to other widely used standardised tests such as the PAT-Reading and PAT-Mathematics tests, but their principal difference is that they are whole-cohort tests taken at the same time across the nation. The stated purposes of Australia's national standardised assessment program have developed over time and have included monitoring progress towards national goals, school system accountability and performance, school improvement, individual student learning achievement and growth, and information for parents/carers on school and student performance.

This section of the report follows the development of these multiple purposes through the Hobart, Adelaide, Melbourne and Alice Springs (Mparntwe) education declarations and national ministerial council decisions.

## **Monitoring national, state and territory programs and policies**

In the Hobart Declaration (Australian Education Council, 1989) Ministers agreed to production of an annual National Report on Schooling that would 'monitor schools' achievements and their progress towards meeting the agreed national goals.' Annual National Reports on Schooling would 'report on the school curriculum, participation and retention rates, student achievements and the application of financial resources in schools'. The subsequent Adelaide Declaration (Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA), 1999) committed Ministers to report on progress towards national goals, including 'explicit and defensible standards ... through which the effectiveness, efficiency and equity of schooling can be measured and evaluated'.

These links between student achievement and national policy were sharpened in 2005 with the ministerial council's agreement to common national standardised tests as 'a means of improving comparability of results among states and territories' (MCEETYA, 2005, p. 2). The links were broadened in the Melbourne Declaration (MCEETYA, 2008a) in which governments committed to the availability of:

Good quality data [that] enables governments to analyse how well schools are performing, identify schools with particular needs, determine where resources are most needed to lift attainment, identify best practice and innovation, conduct national and international comparisons of approaches

and performance and develop a substantive evidence base on what works (p. 17).

The Alice Springs Declaration affirmed that ‘assessment results that are publicly available at the school, sector and jurisdiction level’ provide information that enables ‘policy makers and governments to make informed decisions based on the evidence’ (Education Council, 2019a, p. 11). The declaration went on to list a range of policy domains in which governments rely on ‘good quality data’. Although neither the Melbourne nor Alice Springs declarations specify that the source of such ‘data’ need include whole-cohort national standardised testing, since 2008 annual National Reports on Schooling in Australia have updated progress towards the Melbourne Declaration goals using NAPLAN results.

In the decade between the Melbourne and Adelaide Declarations, ministerial councils have approved a range of developments and uses of whole-cohort test results. In 2009 the ministerial council announced publication of ‘relevant, nationally comparable information on all schools’ using whole-cohort NAPLAN data, on what came to be the My School website (MCEETYA, 2009, p. 1). Subsequent ministerial councils have supported NAPLAN and My School’s role in monitoring government programs and policies. In 2015 the ministerial council reaffirmed its commitment to My School and nationally consistent school level information ‘for the use of parents/carers, school communities and governments’ (Education Council, 2015 p. 2). In 2019, when Ministers agreed to some changes in the representation of NAPLAN data on My School, they noted that ‘gain measures will tell us if students are making the progress they should – and tell us if Australia’s education system is on track.’ (Education Council, 2019c, p. 2).

### **System accountability and performance**

The accountability and performance goals in the initial Hobart Declaration were modest, announcing a commitment to a National Report on Schooling that would ‘increase public awareness of the performance of our schools as well as make schools more accountable to the Australian people.’ The Adelaide Declaration focused on increasing public confidence rather than accountability but committed governments to ‘increasing public confidence in school education through explicit and defensible standards that guide improvement in students’ levels of educational achievement and through which the effectiveness, efficiency and equity of schooling can be measured and evaluated.’ The Melbourne Declaration further committed Australian governments to strengthen accountability and transparency, providing parents/ carers and the community with ‘access to information about the performance of their school compared to schools with similar characteristics’ without resorting to ‘simplistic league tables or rankings’ (p. 17). Similarly, the Alice Springs Declaration committed governments to ‘strengthening accountability and transparency with strong, meaningful measures’. This includes:

assessment results that are publicly available at the school, sector and jurisdiction level to ensure accountability and provide sufficient information to parents, carers, families, the broader community, researchers, policy makers and governments to make informed decisions based on evidence (p. 18).

Between the Melbourne and Alice Springs Declarations, ministerial councils have balanced the goals of accountability and transparency against the risk of misuse of school-level achievement data. The 2009 ministerial council communique that announced the decision to establish My School characterised it as ‘a major step

forward for the shared national transparency agenda' but added the caution that 'Ministers agreed that these reforms were not about simplistic league tables which rank schools according to raw test scores' (MCEETYA, 2009a, p. 2).

Later that year, the ministerial council endorsed a more detailed set of principles for reporting on schooling, affirming that reporting should be in the public interest, use valid, reliable and contextualised data, be sufficiently comprehensive to enable proper interpretation and should balance the community's right to know with the need to avoid misuse of the information (MCEETYA, 2009b, p. 4). The principles also affirmed a variety of purposes for achievement data. Schools need data on the performance of their students because they have the primary accountability for student outcomes. Parents/carers' need information to make informed judgements, make choices and engage with their children's education and the school community. More broadly, the community needs schools to be accountable for the results they achieve with the public funding they receive, and governments are accountable for the decisions they take. Finally, the principles for reporting affirm the need of school systems/sectors and governments for sound information on school performance to support ongoing improvement for students and schools.

Notwithstanding these agreed principles, reporting of national standardised test results at the school level continued to be a matter of concern for Ministers. In a communique following a 2011 meeting, Ministers considered a range of improvements to the My School website but 'reiterated strong opposition to the publication of league tables arising from My School data.' (MCEETYA, 2011, p. 2). In 2018 Ministers 'reiterated their commitment to standard and fair assessment supported

by transparent reporting' (Education Council, 2018, p. 2) and commissioned a review of the reporting on NAPLAN results. In 2019 Ministers noted a number of recommendations designed to reduce the risk of misusing NAPLAN data (Ministerial Council, 2019b). The 2020 version of My School had fewer NAPLAN displays and a clearer focus on school-level estimates of gains in achievement.

### **School improvement**

Neither the Hobart nor the Adelaide declarations linked achievement data and school improvement. The Melbourne Declaration was the first to focus on this theme, referring to the way in which good quality data supports each school in 'the design of high quality learning programs' and 'informs schools' approaches to provision of programs, school policies, pursuit and allocation of resources' (p. 17). Similarly, the Alice Springs Declaration, notes that good quality data:

allows teachers to evaluate the effectiveness of their classroom practice and supports educators to effectively identify learners' progress and growth, and design individualised and adaptive learning programs. It also informs programs, policies, allocation of resources, relationships with parents and partnerships and connections with community and business (p. 18).

In alluding to 'good quality data' neither of these Declarations specify NAPLAN or any other whole-cohort standardised test data. The links between population testing, school-level data and school improvement, however, were addressed more specifically in the Hon. Julia Gillard MP's second reading speech to Parliament supporting the Australian Curriculum, Assessment and Reporting Authority Bill (Australia, Parliament, 2008).

Accurate information on how students and schools are performing tells teachers, principals, parents and governments what needs to be done. This means publishing the performance of individual schools, along with information that puts that data in its proper context. That context includes information about the range of student backgrounds served by a school and its performance when compared against other 'like schools' serving similar student populations.

Australian governments' linkage of NAPLAN scores to school improvement have developed substantially in the last decade. Since 2010, My School has made it possible to compare NAPLAN growth and achievement among schools serving similar students. In recent years school systems/sectors have also made substantial investments in business intelligence systems that (among other things) help schools to explore links between standardised tests results and school improvement.

### **Individual student learning achievement and growth**

The state and territory assessments that preceded NAPLAN provided schools with information about individual student achievement, but these individual results were brought to a common scale with the introduction of NAPLAN. When announcing the first round of NAPLAN tests, Minister Gillard characterised these new tests as having 'a strong diagnostic approach' (MCEETYA, 2008b). In similar terms, the National Report on Schooling in Australia 2008 noted that 'The data from NAPLAN test results gives schools and systems a diagnostic capacity to identify individual student needs' (MCEETYA, 2008c, p. 2) and that 'NAPLAN can be used by teachers for diagnostic purposes' (p. 17). In subsequent years, NAPLAN's move from pen and paper to online testing was expected to provide 'more accurate assessment of each child's

strengths and weaknesses' and 'even greater effectiveness as a diagnostic tool in classrooms' (Education Council 2014, p. 1).

More modest claims about the diagnostic value of NAPLAN at an individual level have been made elsewhere. ACARA's submission to the 2013 Senate inquiry noted that 'NAPLAN tests do not conform to the meaning of 'diagnostic' assessment in the way that this term is commonly understood in a classroom context, as for an individual student there are insufficient items at each difficulty level to provide the detailed information that a diagnostic test is designed to do' (ACARA, 2013b, p. 8). As ACARA's submission went on to say, NAPLAN seeks to complement the assessment tools classroom teachers use by showing how students are performing against national standards. Consistent with this view, the current description of NAPLAN on the National Assessment Program website notes that 'The results can assist teachers by providing additional information to support their professional judgement about students' levels of literacy and numeracy attainment and progress' and 'do not replace the extensive, ongoing assessments made by teachers about each student's performance' (ACARA, 2020).

### **Information for parents/carers on school and student performance**

NAPLAN's national standardised cohort testing has provided two streams of information for parents/carers: individual student reports on NAPLAN growth and achievement and school-level summaries of NAPLAN performance.

The individual reports to parents/carers were anticipated at the launch of NAPLAN in 2008 when Minister Gillard said that the new test program would allow 'parents/carers to understand the level of achievement of students ... [including] information on students who have not achieved the

minimum literacy and numeracy standard and need further support.’ For the last decade ACARA has provided a standard student report that schools and school systems/sectors then pass on for parents and carers of each student who has taken a NAPLAN test. The reports show each student’s performance in achievement bands and compared with the range of achievement for the middle 60% of students taking the same test.

The second stream of data for parents/carers is in the form of school-level summaries of NAPLAN achievement and progress. Launching My School in 2010, Minister Gillard noted that it would allow parents/carers and school communities ‘to compare their school’s results with neighbouring schools and up to 60 statistically similar schools.’ A subsequent ministerial council communique described these as ‘fair comparisons of schools in Australia, letting parents/carers, educators and members of the general public see how a school is performing, compared to schools with similar students’ (Education Council, 2016, p. 1). The statistical basis of the comparisons has changed over time. Initially comparisons were available either with statistically similar schools or students with the same starting point in achievement. In 2020, My School introduced a composite comparison of the amount of improvement achieved by students with the same starting score and a similar socio-educational background. However, following a 2019 Education Council agreement, the website clarified that ‘The inclusion of data about how schools perform in NAPLAN provides information on only one aspect of school performance and does not measure overall school quality’ (ACARA, 2020d).

My School provides NAPLAN achievement and progress scores for every Australian school, but school-level results are available from other sources in many Australian

jurisdictions. The Queensland Curriculum and Assessment Authority website provides a downloadable table of results for public, Catholic and independent schools. School-level results for Western Australian public schools are also available on a searchable website. More commonly, annual reports containing NAPLAN data are available on school websites. In South Australia, for example, annual school reports on government schools’ websites use a standardised format including proportions of students achieving NAPLAN proficiency levels, the proportion of students in three NAPLAN progress bands and the proportion achieving in the upper two bands. In the ACT, annual school board reports on government school websites include a table of NAPLAN mean scores and an annual action plan detailing progress against NAPLAN targets. Victorian public schools provide an annual report to their school community, including NAPLAN results and comparisons with schools with similar characteristics. In NSW, annual reports appear on public school websites but do not publish NAPLAN results in a common format. NSW independent schools are required to publish annual reports that disclose comparative performance over time, comparisons with state-wide performance and comparisons with similar schools where appropriate.

## Summary

### Current purposes of the national standardised testing program

Standardised testing is widespread in Australian schools and has a long history. There are many forms of standardised testing, of which NAPLAN is just one – they are whole-cohort tests of cognitive skills in literacy and numeracy, marked digitally and by experts, reporting against national standards and benchmarks. In the years since the first national declaration in Hobart,

Australia has moved from separate tests in each state and territory to common whole-cohort literacy and numeracy testing. During this time, five purposes for national standardised testing have been endorsed by Australian governments.

**Monitoring.** Removing examinations at the end of primary and junior secondary schooling in most jurisdictions led to the development of state-based literacy and numeracy assessments. Some of these tests were sample tests and others were census tests. To improve the comparability of these tests across jurisdictions, national benchmarks were developed and backed by statistical moderation. These tests were replaced by common, census NAPLAN tests in 2008. Since the first cycle of NAPLAN tests, Australian governments have consistently endorsed the use of NAPLAN's whole-cohort standardised test results for monitoring progress towards national goals.

**Accountability.** Early national declarations committed governments to increasing public confidence in school education. However, since the Melbourne Declaration, Australian governments have reiterated their commitment to transparent school-level reporting of standardised national assessments. Ministers' commitment to accountability has been balanced by their concerns about the possibility of misuse of the data in school rankings and comparisons.

**School improvement.** National, state and territory improvement targets are currently expressed in terms of NAPLAN achievement. School systems/sectors have made substantial investments in business intelligence systems that enable schools and system/sector officials to explore trends in achievements at the individual school level. They have also supported intervention programs designed to help individual schools lift their NAPLAN achievement standards.

**Individual achievement.** Standardised testing can provide teachers with useful information about individual achievement and assist them in planning for students' growth in achievement. Early characterisations of NAPLAN as diagnostic at the individual student level have been moderated in recent years, with ACARA characterising the results as setting the achievements in a national and state or territory context, and triangulating with teachers' judgements and other standardised assessments of student achievement and progress.

**Information for parents/carers.** Individual student results have been available from census NAPLAN testing for more than a decade. Results are provided by ACARA in a standard format that situates individual achievement in the context of all Australian students' achievement and in terms of bands of achievement. In addition, school-level information on NAPLAN growth and achievement has been available on My School since 2010. These forms of information could only be provided in a more limited form if NAPLAN were a sample-based testing program.

### **Stakeholder perspectives on the purposes of national standardised assessment**

Stakeholders responded to the NAPLAN Review in face-to-face and web-conferencing meetings, in a set of Practitioner Reference Group meetings, in formal written submissions and in response to a short survey on the review website. Much of the commentary touched on the purposes of NAPLAN. Although the documentary evidence shows that successive state, territory and national governments have endorsed five purposes for NAPLAN, there is no consensus among stakeholders that these purposes are all equally legitimate – or that



they can be achieved with a single set of standardised assessments.

There was widespread support from school systems/sectors, unions, principals, teachers and parents/carers for using national standardised assessments to monitor national and jurisdictional trends in performance. As one of the teachers' unions said, notwithstanding their fierce criticism of the current NAPLAN program, they have 'an absolute commitment to a national assessment program'.

Support for NAPLAN's use in school-level accountability was much more contested. School system/sector authorities acknowledged 'the accountability regimes in schools and offices are reliant on NAPLAN'. Although their staff 'dislike NAPLAN', they also acknowledge that 'they need it'. Another system official said that despite schools' 'strong opposition' to NAPLAN they 'are willing to accept that NAPLAN is part of accountability'. The broad seam of discontent about NAPLAN and accountability stems primarily from concerns about public comparisons on My School and in the news media. Stakeholders representing a variety of perspectives on school education called for the publication of school-level results to stop. One of the key issues raised was the perception of unfairness of comparisons. As one stakeholder put it:

Take the media out of it to stop unfair comparisons of schools. My school does well in NAPLAN and the school down the road does not do as well. This is not because of poor teaching or student learning; they have a very different clientele who have different base levels of learning. It makes my school look good, but I don't think it is fair on all the other schools (Respondent to the online survey).

The third agreed purpose for NAPLAN is school improvement. School systems/sectors valued the use of NAPLAN data for this purpose. As one said, "NAPLAN has provided us with the opportunity to track schools over time". Another said, "We want school data – allows us to open up conversations with schools." Views of school stakeholders about the value of NAPLAN in school improvement were mixed. One of the school principal representatives characterised NAPLAN as "a lever for innovation in pedagogy and practice." Others described the way in which schools use NAPLAN for identifying weaknesses in test domains or school cohorts, and for forward planning and identifying professional learning goals. Some thought that NAPLAN was valued more by people working at a system/sector and leadership level than by people in schools. Another less positive view expressed by a school principal was that there was just too much measurement going on altogether, "We're on a hamster wheel to prove what we're doing works".

Unions were not opposed, in principle, to using data for school improvement but contested whether assessments designed 'to properly address the needs of teachers, students and families' could also be relied on to target school interventions. There was also some ambivalence about using NAPLAN results to set school targets. As one of the principals' associations argued, 'When a measure becomes a target, it ceases to be a good measure.' Others took the view that NAPLAN was valuable for school improvement as well as system/sector monitoring. As one of the principals' association representatives said:

We believe it can do all those things... Schools want it at the school level – student performance analyser – that would allow us to add the data, teacher

judgement, and the VCE. And that would show us real trends, for instance the year our Year 9 writing improved was the year our median score improved. And then systems need to make sure policy is working... how are we going to assess whether it was worth it and whether [support] was going to the right place?

Regarding NAPLAN's value in supporting individual improvement, the most common concern was about the balance between standardised assessment scores and teachers' judgements about students' achievement. Although a few submissions rejected the use of standardised assessment entirely, more commonly the issue was characterised in terms such as these, 'over-reliance on standardised testing... diminishes teacher professional judgement' (Member of the NAPLAN Review Practitioners' Reference Group). NAPLAN tests were seen as just one source of data, to be used in triangulation with other information. In addition to views about the primacy of teacher judgment, some stakeholders would prefer assessment that provide more fine-grained diagnostic information than is available from NAPLAN. There is a clear call, one of the union submissions argued, 'from classroom practitioners and principals for an assessment program that is more closely linked to what teachers do in their classrooms and what would assist in diagnosing student strengths and weaknesses'. Many of the teachers who responded took this view, arguing that NAPLAN needs to:

be designed in a way that the main purpose is to identify and address weaknesses in national, community and individual key skills. Be a component of the real teaching year and generate class and individual lesson plans based on what was found in the NAPLAN results. Allow

monitoring of individuals and classes over time to see improvement in weaknesses. (Respondent to the online survey)

The final purpose that governments have set out for NAPLAN over time is to provide information for parents/carers. Although not all parent stakeholder groups were enthusiastic about NAPLAN, and especially the comparisons that could be made among schools using My School, there was broad recognition of the right of parents/carers to be informed about students' achievement and some acknowledgement that standardised testing could provide information that is independent of the local context. As one submission from a parents/carers' group put it,

[S]tandardised assessment is important because it provides an independent benchmarked measure as part of a well-rounded assessment of each child's learning achievement and growth. It is particularly important for children in smaller schools and rural or remote areas, where the opportunities for comparisons may be limited. While the context and detailed knowledge gained by teachers in their day to day professional dealings with individual students cannot be replaced, and will always be valuable to students, parents/carers and schools, external independent assessment over time is also useful information for parents/carers and students. Its role is to provide an independent context to balance other information about student learning achievement and growth. (Written submission response: Parents/carers' association)

In addition to commentary on the value of particular purposes for national standardised assessment, there was a good deal of concern that the problem with NAPLAN's purposes was lack of clarity. Some stakeholders suggested that the purposes have 'strayed' from those originally agreed; others expressed this more kindly as the purposes having 'evolved'. One of the school system/sector officials suggested that NAPLAN was being used for things it was not designed for, arguing that, 'It was designed for system level data' but was being used 'for individual student data'. Other stakeholders drew attention to conflicts between some of the current purposes of NAPLAN. As a representative from one of the subject associations said,

'It is difficult to simultaneously achieve census and system testing in conjunction with diagnostic testing for teachers'. To this end, union and principals' association representatives often argued that the accountability and system performance purposes of national standardised testing could be met by sample testing, and that teachers' judgements should be supported by richer, on-demand diagnostic assessments.

# Chapter 2: National and international measures of achievement

This chapter explores the evidence of achievement from more than a decade of NAPLAN census testing and considers that evidence of achievement in the context of achievement arising from three international sample testing programs – the Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA). The final section of the chapter considers this evidence of continuity and change in achievement alongside the feedback on national and jurisdictional performance received during consultation for the NAPLAN Review.

## Key points:

- National NAPLAN results have improved in the last decade in Years 3 and 5 but not in Years 7 and 9. Writing achievement has been static in Years 3 and 5 and has declined in Years 7 and 9.
- Some jurisdictions, notably Queensland and Western Australia, have improved more than others on national and international measures.
- Australia is a middle-ranking country in international test comparisons, ranked below high-performing Asian countries and often below Canada, England and Ireland.
- There are important differences among the national and international tests, including test domains, item types, degree of focus on curriculum content, cognitive demand, number of proficiency levels reported and whether they are sample tests.
- There have been improvements in NAPLAN reading and numeracy in Years 3 and 5, PIRLS Year 4 reading and TIMSS Year 4 mathematics. Although achievement in Years 7 and 9 NAPLAN reading and numeracy has not changed in a decade, PISA reading literacy and PISA mathematical literacy of 15-year-olds have declined.
- The proportion of high-performing students in PIRLS Year 4 reading and TIMSS Year 4 mathematics has increased, but there has been no change in the proportion of high or low performers in TIMSS Year 8. In PISA, the proportion of low-performing students has increased in reading and mathematics and the proportion of high-performing students has decreased in mathematics.
- There is widespread support for a national standardised testing program, as well as widespread concern about limitations of the current NAPLAN program.

## NAPLAN achievement and improvement, 2008 to 2019

Twelve National Reports have been produced since 2008, providing NAPLAN data in the five test domains: reading, writing, spelling, grammar and punctuation and numeracy. Results are also available on an interactive website provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). The most recent 2019 National Report provides mean scale scores, the proportion of students in each of ten achievement bands and the proportion of students above the national minimum standard (NMS). These results are disaggregated by sex, Indigenous status, language background other than English (LBOTE) status, geolocation, parental education and parental occupation.<sup>1</sup> Results in each of these categories are available for Australia as a whole and for each state and territory. Comparisons of results across jurisdictions are reported for all five test domains. In addition to reporting NMS proportions and mean scores in each domain, ACARA provides estimates of the statistical significance of differences

in achievement. When differences are statistically significant and have an effect size between 0.2 and 0.5 they are described as 'moderate'; when they are significant and have an effect size greater than 0.5 they are described as 'substantial' (ACARA, 2019c p. 300).

### Reading

NAPLAN reading tests are designed to 'assess students' ability to read and view texts to identify, analyse and evaluate information and ideas' (ACARA, 2017, p. 9). The reading tests use written texts including some with graphics and images. ACARA notes that much of the teaching of literacy occurs in the English learning area and that the tests are aligned with the literacy aspects of the Australian Curriculum (p. 5).

Australia-wide, there have been moderate improvements between 2008 and 2019 in national mean scale scores and in the proportion of students equal to or above the NMS in Years 3 and 5 reading, but no significant national increases in Years 7 and 9 reading achievement since NAPLAN began (see Table 2).<sup>2</sup>

**Table 2:** Differences in achievements of students in reading, 2008 to 2019

Students		AUS	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
<b>Year 9</b>	Mean	•	•	•	•	△	•	•	•	•
	≥NMS*	•	•	▽	•	•	•	▽	▽	•
<b>Year 7</b>	Mean	•	•	•	△	△	•	•	•	•
	≥NMS	•	•	•	△	•	•	•	•	•
<b>Year 5</b>	Mean	△	•	△	△	△	△	△	△	•
	≥NMS	△	•	△	▲	△	△	△	•	•
<b>Year 3</b>	Mean	△	△	△	▲	△	△	△	△	△
	≥NMS	△	△	△	▲	▲	△	•	•	△

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

1 The NAPLAN *National Reports* use of the term 'Indigenous' to refer to Aboriginal and Torres Strait Islander peoples. In this report, this term Indigenous is used when the specific reference is to data from the *National Reports*; elsewhere, the term Aboriginal and Torres Strait Islander peoples is preferred.

2 Data for these and the following NAPLAN comparison tables are drawn from time series data available at <https://reports.acara.edu.au/Home/TimeSeries>

National improvements in Years 3 and 5 reading were shared across a range of student demographic groups. Both male and female students showed moderate increases in the proportion of Years 3 and 5 students at or above the reading NMS and mean scores between 2008 and 2019. Moderate increases were recorded for Indigenous and LBOTE students in NMS and mean scores in Years 3 and 5 and in mean scores in Year 7. There were no significant differences in Year 9 reading for any of these demographic groups.

The differences among states and territories in mean score achievement that were evident in the first round of testing in 2008 have continued throughout the years of NAPLAN testing. In both 2008 and in 2019, the ACT, Victoria and NSW had the highest mean scores and the Northern Territory had the lowest mean scores in Year 3 reading. The rank order of other jurisdictions changed slightly over time. Queensland's and Western Australia's Year 3 reading mean scores were lower than South Australia's and Tasmania's in 2008 and higher in 2019. In 2019, Victoria's Year 3 reading mean score was statistically similar to NSW and the ACT, and superior to all other jurisdictions. Similar patterns occurred in Years 5, 7 and 9 reading in both 2008 and 2019, where the ACT, Victoria and NSW were typically the three highest performing jurisdictions and the Northern Territory was the lowest performing jurisdiction. Among the other states, moderate improvement in Western Australia's and Queensland's mean scores improved their rank order compared with South Australia and Tasmania.

In Queensland, there were substantial increases in Year 3 NMS and mean scores, substantial increases in Year 5 NMS and moderate increases Years 5 and 7 NMS and mean scores. Changes in Western Australia included a substantial increase in the percentage of Year 3 students at or

above the NMS and moderate increases in NMS and mean scores in Years 3, 5, 7 and 9 reading.

Between 2008 and 2019, there were moderate increases in NMS and mean scores in Years 3 and 5 reading in Victoria and South Australia, in Year 3 NMS and mean scores in the Northern Territory and in Year 3 mean and Year 5 NMS and mean scores in Tasmania. In NSW, there were moderate increases in NMS and mean scores in Year 3, but no changes in the NMS or mean scores in other Years. The ACT showed moderate increases in Years 3 and 5 mean scores, but no change in the proportion of students at or above the NMS. In Victoria, Tasmania and the ACT there were decreases in the proportion of students at or above the NMS in Year 9 reading, but no change in the mean scores from 2008 to 2019.

In sum, the evidence of a decade's NAPLAN results is that there have been moderate improvements in reading achievement nationally and in most jurisdictions in Years 3 and 5. Although there has been no national improvement in Years 7 and 9 reading scores, there have been moderate declines in the proportion of students achieving at or above the NMS in several jurisdictions and improvements in that proportion and in mean scores in other jurisdictions. The performance of two states stands out: Western Australia's moderate improvements in mean scores in Years 3, 5, 7 and 9, and Queensland's moderate improvement in Years 5 and 7 and substantial improvement in Year 3 reading achievement, though both remain behind the ACT, NSW and Victoria.

## Writing

The NAPLAN writing tests are aligned to the Australian Curriculum in English through a focus on seven sub-strand threads of the curriculum – purpose, audience and structures of different types of texts, vocabulary, text cohesion, sentences and

clause level grammar, word level grammar, punctuation and spelling (ACARA 2017, p.14).

The assessment of writing has been subject to greater turbulence than the other NAPLAN measures and is discussed in detail in Chapter 5. One of the sources of turbulence has been the possibility that the writing prompts in any one year could be for either narrative or persuasive writing, leading initially to two separate assessment scales, which made year-to-year comparisons difficult. Since 2011, however, the prompts have been for persuasive writing in every year except 2016. NAPLAN narrative writing scores from 2016 have since been mapped onto the existing writing scale, providing results and trend data from 2011 to 2019 on a single scale. In the comparison below, 2011 is the base year for writing (see Table 3).

There has been little change in either the national NMS proportions or mean scores in writing over nine years, and what little change there has been is negative. Between 2011 and 2019 there were no national improvements in either NMS or mean scores, and there were moderate declines in mean scores in Years 7 and 9. Indigenous students achieved a moderate increase in NMS proportions.

Among the states and territories there were moderate increases in mean scores in Year 3 in Western Australia and Tasmania and moderate increases in the proportion of students at or above NMS in Year 3 in Western Australia and Queensland. In some other jurisdictions there were declines in performance between 2011 and 2019. Mean scores in the ACT decreased in Years 5, 7 and 9. In Queensland, NMS and mean scores declined in Years 7 and 9.

**Table 3:** Differences in achievements of students in writing, 2011 to 2019

Students		AUS	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
Year 9	Mean	▽	.	▽	▽	.	.	.	▽	▽
	≥NMS*	.	.	.	▽	.	.	.	.	.
Year 7	Mean	▽	.	.	▽	.	▽	.	▽	.
	≥NMS	.	.	.	▽	.	.	.	.	.
Year 5	Mean	.	▽	.	.	.	.	.	▽	.
	≥NMS	.	.	.	.	.	.	.	.	.
Year 3	Mean	.	.	.	.	△	.	△	.	.
	≥NMS	.	.	.	△	△	.	.	.	.

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

## Spelling

Spelling is one of two parts of the NAPLAN language conventions tests. The second part is grammar and punctuation. The spelling tests focus on the accurate spelling of written words and draw on the Australian Curriculum English sub-strand of spelling. Students are required to either provide the correct spelling of a designated word or identify a misspelled word and then write the correct spelling (ACARA, 2017, p. 13).

There is some NAPLAN evidence of improvement in national spelling achievement since 2008, with moderate increases in mean scores in Year 3 and NMS and mean scores in Year 5 (see Table 4). As has been the case in other test domains, there is no evidence of changes in national achievement in the secondary school test years, Years 7 and 9.

Among the demographic groups, there have been moderate increases in mean spelling scores of Year 3 males and Years 3 and 5 males and females, as well as moderate increases in NMS and mean

scores of Indigenous students in Years 3 and 5. Among LBOTE students there have been moderate increases in mean scores for Year 3 and moderate NMS and mean score increases for Year 5.

The modest national improvements in spelling achievement appear to have been driven by improvements in two states – Queensland and Western Australia. From 2008 to 2019, there was a substantial increase in Queensland's Year 3 mean scores and moderate increases were observed in Years 5 and 7 mean scores. Moderate increases in the proportion of Year 3 students at or above the NMS were recorded in Years 3, 5, 7 and 9. In Western Australia, moderate increases were achieved in both NMS and mean scores in Years 3, 5, 7 and 9. In contrast, there were no other states or territories that recorded improvements in spelling scores by either measure; and NSW and Tasmania recorded a moderate decrease in the proportion of students at or above the Year 3 NMS.

**Table 4:** Differences in achievements of students in spelling, 2008 to 2019

Students		AUS	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
<b>Year 9</b>	Mean	•	•	•	•	△	•	•	•	•
	≥NMS*	•	•	•	△	△	•	•	•	•
<b>Year 7</b>	Mean	•	•	•	△	△	•	•	•	•
	≥NMS	•	•	•	△	△	•	•	•	•
<b>Year 5</b>	Mean	△	•	•	△	△	•	•	•	•
	≥NMS	△	•	•	△	△	•	•	•	•
<b>Year 3</b>	Mean	△	•	•	▲	△	•	•	•	•
	≥NMS	•	▽	•	△	△	•	▽	•	•

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease



## Grammar and punctuation

The second part of the language conventions test concerns grammar and punctuation. The grammar items focus on knowledge and accurate use of grammar at a sentence, clause and word level. Grammar items are developed from the sub-strand threads of text cohesion, sentences and clause level grammar and word level grammar in the Australian Curriculum in English. The punctuation items are developed from the content of the sub-strand thread of punctuation (ACARA 2017, p. 12).

Evidence of national improvement in scores on the NAPLAN grammar and punctuation tests is scant (see Table 5). In Year 3 alone, there were moderate increases in national mean scores and in the proportion of students at or above the NMS.

Mean and NMS scores increased for both male and female Year 3 students between 2008 and 2019. Indigenous students' and LBOTE students' achievement improved on both measures in Year 3, and their mean scores increased in Year 7.

Across the states and territories, there were moderate mean score increases in Year 3 grammar and punctuation in NSW, South Australia, the ACT and the Northern Territory and substantial mean score increases in Queensland and Western Australia. Although there were no changes in Years 5, 7 or 9 scores on either measure in the other states and territories, there were moderate increases in Western Australia's NMS and mean scores in Year 7 and 9 and Queensland's mean score in Year 7.

**Table 5:** Differences in achievements of students in grammar and punctuation, 2008 to 2019

Students		AUS	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
<b>Year 9</b>	Mean	•	•	•	•	△	•	•	•	•
	≥NMS*	•	•	•	•	△	•	•	•	•
<b>Year 7</b>	Mean	•	•	•	△	△	•	•	•	•
	≥NMS	•	•	•	•	△	•	•	•	•
<b>Year 5</b>	Mean	•	•	•	•	•	•	•	•	•
	≥NMS	•	•	•	△	•	•	•	•	•
<b>Year 3</b>	Mean	△	△	•	▲	▲	△	•	△	△
	≥NMS	△	•	•	▲	△	•	•	•	•

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

## Numeracy

The NAPLAN numeracy tests assess students' application of mathematical knowledge and skills in everyday contexts. Test items draw on both the content and proficiency strands of the Australian Curriculum in mathematics. The proficiency strands are understanding, fluency, problem solving and reasoning. The content strands are number and algebra, measurement and geometry and statistics and probability (ACARA 2017, p. 17). The Years 7 and 9 numeracy tests include some items for which calculators may be used and some non-calculator items. Calculators are not used for Years 3 or 5 tests.

Unlike the Year 3 NAPLAN literacy tests, where there were moderate national mean score improvements in achievement in the last decade, there is no evidence of national improvement in the Year 3 numeracy achievement over that time (Table 6). There were moderate increases in NMS and mean scores in Year 5, however, and an increase

in the proportion of students at or above the NMS in Year 7 numeracy. Among the demographic groups, increases in Year 5 NMS proportions and mean scores and the NMS proportion in Year 9 numeracy were achieved by males and females, Indigenous students and LBOTE students.

Across the jurisdictions, the pattern of differences in achievement from 2008 to 2019 was inconsistent. Although there were national increases in Year 5 numeracy, mean scores and NMS proportions did not change in NSW or the Northern Territory. Queensland recorded substantial increases in mean scores and NMS proportions, while Victoria, Western Australia and South Australia recorded moderate increases on both Year 5 numeracy measures. In Year 9, Queensland recorded moderate increases on both measures and Western Australia recorded a substantial increase in NMS proportions and a moderate increase in mean scores.

**Table 6:** Differences in achievements of students in numeracy, 2008 to 2019

Students		AUS	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
<b>Year 9</b>	Mean	•	•	•	△	△	•	•	•	•
	≥NMS*	△	△	•	△	▲	△	△	•	•
<b>Year 7</b>	Mean	•	•	•	•	△	•	•	•	•
	≥NMS	•	▽	•	•	•	•	▽	•	•
<b>Year 5</b>	Mean	△	•	△	▲	△	△	△	△	•
	≥NMS	△	•	△	▲	△	△	•	•	•
<b>Year 3</b>	Mean	•	•	•	△	△	•	•	•	•
	≥NMS	•	•	•	△	•	•	•	•	•

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

## Patterns of change across the NAPLAN test domains

Looking across the test domains, evidence of national improvement in NAPLAN achievement is overwhelmingly in the primary school years (Table 7). There have been moderate increases in Years 3 and 5 mean scores in reading, spelling and grammar and punctuation; and NMS increases in both reading, and grammar and punctuation. There is no evidence of national improvement in literacy in either Years 7 or 9, and evidence of moderate declines in writing in both Years 7 and 9.

In the secondary years, the only evidence of moderate improvement is in the proportion of students meeting the NMS for Year 9 numeracy. There are no significant improvements in mean scores in any of the Year 7 or 9 test domains, and mean scores showed moderate declines in both Years 7 and 9 writing.

**Table 7:** Differences in achievement, base year to 2019, by test domain, Australia

Students		Reading	Writing	Spelling	Grammar & punctuation	Numeracy
<b>Year 9</b>	Mean	.	▽	.	.	.
	≥NMS*	.	.	.	.	△
<b>Year 7</b>	Mean	.	▽	.	.	.
	≥NMS	.	.	.	.	.
<b>Year 5</b>	Mean	△	.	△	.	△
	≥NMS	△	.	△	.	△
<b>Year 3</b>	Mean	△	.	△	△	.
	≥NMS	△	.	.	△	.

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

Among the demographic groups tracked in the NAPLAN National Reports, there is evidence of moderate improvement by Indigenous students on a range of measures. Although Indigenous students' NAPLAN results continue to lag behind those of non-Indigenous students, Indigenous students' mean scores or NMS proportions increased in Year 3 reading, writing, spelling and grammar and punctuation; Year 5 reading, spelling and numeracy; Year 7 reading and grammar and punctuation; and Year 9 numeracy.

The sense that there has been limited national improvement over the last decade may be contrasted with the experience of individual states and territories. Across the five test domains and four test year levels, Queensland produced moderate or substantial increases in 11 mean scores and 13 NMS proportions. Similarly, Western Australia increased mean scores in Years 7 and 9 in all but the writing tests, which held steady against a national trend of reduced achievement in Years 7 and 9 writing. By way of example, differences in achievement in each test domain and test year from the base year to 2019 for Western Australia are reproduced in Table 8.

**Table 8:** Differences in achievement, base year to 2019, by test domain, Western Australia

Students		Reading	Writing	Spelling	Grammar & punctuation	Numeracy
Year 9	Mean	△	.	△	△	△
	≥NMS*	.	.	△	△	▲
Year 7	Mean	△	.	△	△	△
	≥NMS	.	.	△	△	.
Year 5	Mean	△	.	△	.	△
	≥NMS	△	.	△	.	△
Year 3	Mean	△	△	△	▲	△
	≥NMS	▲	△	△	△	△

\* NMS: National minimum standard

**Key:** • No change △ Moderate increase ▲ Substantial increase ▽ Moderate decrease ▼ Substantial decrease

## NAPLAN and international surveys of student achievement

Three international surveys taken by students in Australian schools overlap the knowledge and skills tested by NAPLAN. The Progress in International Reading Literacy Study (PIRLS) assesses Year 4 students on a five-year cycle. Australia has participated in the 2011 and 2016 cycles. The Trends in International Mathematics and Science

Study (TIMSS) assesses students in Years 4 and 8 in mathematics and science. Tests are on a four-year cycle. Since 1995, Australia has participated in six cycles. The Programme for International Student Assessment (PISA) assesses 15-year-old students' ability to use their science, reading and mathematics knowledge and skills to meet real-life challenges. Assessments are on three-year cycles and Australia has participated in PISA reading, mathematics and science since 2000.

## PIRLS and NAPLAN reading

The PIRLS reading assessment framework identifies two purposes for reading that account for most reading activities done by young students inside and outside the classroom, for literary experience and to acquire and use information (Thomson et al., 2017a). Four processes of comprehension are assessed within each of these two major reading purposes:

- focusing on and retrieving explicitly stated information
- making straightforward inferences
- interpreting and integrating ideas and information
- examining and evaluating content, language and textual elements.

The PIRLS reading purposes and processes have a substantial overlap with the Year 4 Australian Curriculum in English.<sup>3</sup> These purposes and processes are also broadly consistent with the NAPLAN reading tests which focus on assessing ‘students’ ability to read and view texts to identify, analyse and evaluate information and ideas’ (ACARA, 2017, p. 9).

Fifty countries participated in PIRLS 2016 (Thomson et al, 2017a). Australia was in the middle achievement group, with a higher achievement than 24 countries, similar achievement to 13 countries and lower achievement than 13 countries. Higher performing countries included most of the Asian countries, Ireland and England. Similar performing countries included the United States and Canada and lower performing countries included France and New Zealand.

Australia recorded a significant improvement in the average reading score between the 2011 and 2016 PIRLS

assessments, consistent with the statistically significant improvements in reading achievement in NAPLAN Years 3 and 5 mean scores. Among the highlights of this improvement were students’ relative strength in the literary reading purpose without a relative weakness in acquiring and using information. Regarding the processes, Australian Year 4 students had a relative strength in the interpreting, integrating and evaluating processes scale, and a relative weakness in the retrieving and straightforward inferencing scale.

From PIRLS 2011 to PIRLS 2016 there were also some differences in patterns of improvement across states and territories:

- The performance of students in Victoria was significantly higher than that of students in all other jurisdictions except the ACT.
- Students in South Australia performed significantly lower, on average, than students in Victoria, the ACT and Western Australia.
- Western Australia showed the greatest improvement of 28 points from PIRLS 2011 to 2016, followed by Queensland (26 points) and Victoria (21 points). There was no significant change in average scores between 2011 and 2016 in the remaining jurisdictions.

These inter-jurisdictional differences are broadly consistent with the evidence of change in NAPLAN Years 3 and 5 reading over the last decade. Victoria and the ACT (along with NSW) have typically had the highest scores in the country. Queensland and Western Australia have had the strongest patterns of improvement, and their rank order improvement has been at the expense of South Australia.

<sup>3</sup> These descriptions of the three international surveys and their relationship to the NAPLAN tests draws heavily on an advice paper prepared for the NAPLAN Review by Dr. Sue Thomson of the Australian Council for Educational Research, May 2020.

## TIMSS and NAPLAN numeracy

The TIMSS assessment frameworks in numeracy are organised around two dimensions (Grønmo, Lindquist, Arora, & Mullis, 2013). The content dimension specifies the subject matter to be assessed and the cognitive dimension specifies the thinking processes to be assessed. The content dimensions are different for the Years 4 and 8 tests, reflecting content commonly taught at each year level. The Year 4 TIMSS content domains are number, geometric shapes and data display; the Year 8 content domains are number, algebra, geometry and data and chance. There is a greater concentration of test content in number in Year 4; and in Year 8 the content focuses more on interpreting data and the fundamentals of probability. There are three TIMSS cognitive domains – knowing, applying and reasoning. In Year 4, TIMSS has less emphasis on the knowing domain and greater emphasis on the reasoning domain than in Year 8.

According to an advice paper prepared for the NAPLAN Review, the content domains of TIMSS are very similar to the Australian Curriculum in mathematics, which in turn underpins the NAPLAN numeracy tests. There are, however, some differences of emphasis in the cognitive domain. The Australian Curriculum emphasis on knowing and applying is similar to TIMSS but the Australian Curriculum does not appear to cover some of the complexity that is described in the TIMSS framework under reasoning. It seems likely, too, that a substantial number of TIMSS mathematics items are beyond Australian Curriculum expectations for achievement, especially at the Year 4 level.

Fifty-seven countries participated in TIMSS 2015 (Thomson et al., 2017b). In Year 4 mathematics, Australian students' mean score was higher than 20 countries, lower

than 21 countries and not different from seven countries. Countries with higher scores included the TIMSS high-performing Asian countries, Ireland, England and the United States; countries with similar scores included Canada and Germany; and countries with lower scores included Italy, Spain and New Zealand. In Year 8 mathematics, 12 countries had higher performances than Australia and 21 had lower performances. Higher performing countries included TIMSS Asian countries, Canada, Ireland and the United States; lower performing countries included Italy and New Zealand.

In the TIMSS content dimensions, in 2015 Australian Year 4 students performed better in data display and geometric shapes and measures but were weaker in number. In the cognitive dimensions, they were better in applying and reasoning but were weaker in knowing. Year 8 students performed better in data, and chance and number than in algebra and geometry and were weaker in applying and stronger in reasoning.

Australia's Year 4 score in TIMSS 2015 was a significant improvement on the 1995 national score, but this was due to a single increase recorded in 2007 with no change recorded in 2011 or 2015. For Year 8, Australia's result dipped in 2007 and this was followed by a recovery in 2011. Australia's 2015 Year 8 mathematics score was not significantly different from the corresponding score in 1995.

Although there was no national change in Year 4 TIMSS mathematics between 2011 and 2015, NSW, Queensland, South Australia, Western Australia and Tasmania all had significantly higher average scores in 2015 than in 1995 and Western Australia showed the greatest score improvement. Year 8 mathematics achievement improved in Victoria from 1995 to 2015 and from 2003 to 2015. NSW students' achievement declined

from 2003 to 2015. Western Australia improved from 2003 to 2015 but has not yet returned to its 1995 level of achievement in TIMSS mathematics.

Change in TIMSS mean scores over the long term is broadly consistent with the NAPLAN evidence. TIMSS Year 4 achievement has improved from 1995 to 2015 and so has Years 3 and 5 NAPLAN numeracy from 2008 to 2019; TIMSS Year 8 has been static and so has Years 7 and 9 NAPLAN numeracy over those time intervals.

### **PISA reading literacy and NAPLAN reading**

PISA reading literacy measures the capacity to understand, use and reflect on written texts to achieve goals, develop knowledge and potential, and participate in society (OECD, 2020). Unlike NAPLAN, PIRLS and TIMSS, PISA is not a curriculum-based assessment. The conceptualisation of reading literacy has been revised each time that it has been the major assessment PISA domain (2000, 2009, 2018). The current 2018 reading literacy framework 'integrates reading in a traditional sense together with the new forms of reading that have emerged over the past decades and that continue to emerge due to the spread of digital devices and digital texts' (OECD 2019a, p. 22).

An advice paper prepared for this review noted that there has been no comprehensive comparison between the knowledge and skills assessed by PISA and the levels of achievement in the Year 9 Australian Curriculum, but that the level of the Australian Curriculum may be less advanced than the PISA framework suggests. Balanced against this, however, it is important to note that most Australian students have taken PISA tests in Year 10.

Average performance of Australian students in PISA 2018 reading literacy was higher than students in 58 countries and lower than

those in 10 countries (Thomson et al., 2019). Those that outperformed Australia included the PISA high-performing Asian countries, Canada and Ireland; countries with similar achievement included New Zealand, the United States and the United Kingdom. Countries that Australia outperformed included France, the Netherlands and Italy. There were statistically significant declines in PISA reading literacy from 2000 to 2018 as well as between 2003 and 2006, 2009 and 2015, and 2012 and 2018, but Australian students' achievement did not change from 2015 to 2018. These declines in PISA scores are not reflected in the more curriculum-based NAPLAN Years 7 and 9 reading assessments, which registered no significant difference from 2008 to 2019.

In 2018, the average performance of students from the ACT was higher than that of students in any of the other jurisdictions. The next group of states was Western Australia, Victoria and Queensland. Students in South Australia, NSW and Tasmania performed at a similar but lower level and the lowest performing jurisdictions were Tasmania and the Northern Territory.

### **PISA mathematical literacy and NAPLAN numeracy**

PISA mathematical literacy measures 'capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgements and decisions needed by constructive, engaged and reflective citizens' (OECD, 2019a, p. 75).

Australia outperformed 47 countries, was similar to those in eight countries and lower than those in 23 countries (Thomson et al., 2019). The high-performing Asian countries,

Canada, the United Kingdom and Ireland were among those with higher mean scores; France and New Zealand had similar scores and the United States had lower scores.

Although Australia remained in the middle group of countries in PISA 2018 mathematical literacy, Australia's mean performance in PISA mathematical literacy declined from 2003 to 2018 and in some of the intervals in between (2006 to 2012, 2009 to 2012 and 2012 to 2015), but did not change from 2015 to 2018 (Thomson et al, 2019). This record of long-term decline in PISA mathematical literacy is not matched by the more curriculum-based NAPLAN numeracy tests, which showed no significant changes between 2008 and 2019 in Year 9.

As they have done in the most recent cycles of international assessment and most NAPLAN assessments, students in the ACT performed at a higher level than other jurisdictions in PISA 2018 mathematical literacy. Students in Western Australia and Victoria performed at similar levels. Students in Queensland, NSW and South Australia performed at similar but lower levels, and students in Tasmania and the Northern Territory were outperformed by students in all other jurisdictions. There were, however, declines in all Australian jurisdictions from 2003 to 2018. The largest decline was in South Australia, where the decline was almost equivalent to two years of schooling.

### **The national and international standardised testing programs compared**

There are some differences between the international standardised tests and NAPLAN. PIRLS, TIMSS and PISA are sample tests rather than whole-population tests. PIRLS and TIMSS are both curriculum-based tests and are relatively well-linked to the Australian Curriculum, which underpins NAPLAN tests. PISA focuses on the somewhat different constructs of

reading literacy and mathematical literacy. PISA tests are not curriculum-based and are moving rapidly to embrace testing the uses of literacy and numeracy in digital contexts. A further difference between the Australian and international tests is that PIRLS, TIMSS and PISA all have the capacity to report scores on sub-scales. In PISA, for example, as well as national estimates of overall achievement in reading literacy, performance is reported in terms of the three cognitive subscales (locating information, understanding and evaluating, and reflecting), and two text structure subscales (single-source text and multiple sources texts). It also seems likely that several of the international tests are pitched a little higher in cognitive terms than the corresponding Australian curriculum for that year level.

There are also some differences in item types. TIMSS, PIRLS and PISA include some open constructed response items that require trained markers rather than digital marking. NAPLAN paper tests, other than writing, are now digitally marked. While machine marking strategies may tend to narrow the breadth of skills to be assessed, they do have the advantage of allowing for almost immediate feedback to students and teachers.

Despite the differences in test domains, item types and curriculum focus, at the highest level of generality there are common conclusions to be drawn about Australia's performance and improvement from the international tests. On all three of PIRLS, TIMSS and PISA, Australia is a middle-ranking country. Average performance is below the high-performing Asian countries and often below comparable English-speaking jurisdictions such as Canada, the United Kingdom and Ireland.

Although the time intervals are different (depending on when Australia entered the particular test series and the number of



years between PISA, TIMSS and PIRLS test cycles), there are some broad similarities in performance across the testing programs. Statistically significant improvements in NAPLAN at Years 3 and 5 levels are echoed in improvements in TIMSS Year 4 mathematics and PIRLS Year 4 reading. Flat performance in NAPLAN Years 7 and 9 numeracy is echoed in flat performance in TIMSS Year 8. The one exception is in NAPLAN Years 7 and 9 literacy, where flat performance from 2008 to 2019 corresponds to statistically significant declines in PISA reading literacy from 2000 to 2018. These statistically significant changes are illustrated in Table 9.

The proportion of students at various proficiency levels provides another perspective on comparative performance. PIRLS and TIMSS have low, intermediate, high and advanced benchmarks; PISA reports on proficiency levels 1 to 6, classifying students below level 2 as low performers. The patterns of change in proficiency are similar to those for average achievement across the international assessments.

- In PIRLS, the proportion of Australian students who performed at or above the advanced benchmark increased from

2011 to 2016, and nationally there was no change in the proportion who failed to reach the low benchmark (Thomson et al. 2017a, p. 10).

- In TIMSS Year 4 mathematics the proportion of Australian students achieving at or above the advanced benchmark increased and the proportion of students not achieving the low benchmark decreased in most jurisdictions from 1995 to 2015 (Thomson et al, 2017b, p. 21).
- In TIMSS Year 8 mathematics there was no change in the national proportion of students achieving at or above the advanced benchmark or not achieving the low benchmark from 1995 to 2015 (Thomson et al, 2017b, p. 53).
- In PISA reading literacy, the proportion of low-performing students increased, and the proportion of high-performing students did not change between 2000 and 2018 (Thomson et al 2019, p. 47). In PISA mathematical literacy, the proportion of low-performing students increased, and the proportion of high-performing students decreased between 2003 and 2018 (p. 127).

**Table 9:** Differences in achievement, NAPLAN, PIRLS, TIMSS and PISA

Assessment	Year/age	Interval	Reading/literacy	Mathematics/ numeracy
PISA	15-year-old	2000-2018	▽	N/A
PISA	15-year-old	2003-2018	N/A	▽
NAPLAN	Year 9	2008-2019	•	•
TIMSS	Year 8	1995-2015	N/A	•
NAPLAN	Year 7	2008-2019	•	•
NAPLAN	Year 5	2008-2019	△	△
TIMSS	Year 4	1995-2015	N/A	△
PIRLS	Year 4	2011-2016	△	N/A
NAPLAN	Year 3	2008-2019	△	△

**Key:** • No change △ Significant increase ▽ Substantial decrease N/A not applicable

It is unclear why most of the improvements registered by these national and international assessments have occurred in the primary school years. It may be the result of strengthened national, state and territory efforts to ensure that all students make the best start possible in schools through curriculum and professional development reforms focused on the early years. Alternatively, or additionally, it may reflect a clearer connection between the curriculum and the test domains in primary schools, where most classroom teachers teach both mathematics and English. This is in comparison with the diffusion of responsibility in secondary schools, where literacy and numeracy are important in many school subjects but are more explicitly the responsibility of English teachers and mathematics teachers. Which of these is correct, or whatever other causes there may be, there are multiple sources of evidence that Australia's secondary schools have not shared their primary school colleagues' success in improving achievement in reading/literacy or mathematics/numeracy.

Where there are differences between national and international estimates of achievement, such as the decline in PISA 15-year-old scores and the absence of change in Year 9 NAPLAN reading and numeracy, it would be useful to be able to investigate whether this has been due to differences in test domains or item design by investigating the performance of common students taking both kinds of tests.

## Stakeholder views on national and international standardised testing

Among the stakeholders consulted in this review, it is fair to say that there is not a lot of affection for NAPLAN in its current form. The stakeholder comments summarised in Chapter 1 demonstrated a wide range of views about what NAPLAN's purposes have been, should have been or could be. There was, however, relatively little commentary on what has been learned from a decade of national standardised assessment.

Among the individuals who chose to respond to the review's online survey, an overwhelming majority were critical of NAPLAN, expressing concerns about distortion of teaching programs, lack of diagnostic value, or misuse of results in ranking and comparing schools.

The views of organisational stakeholders were more mixed. Some, like one of the principals' associations, took the view that we "wouldn't lose anything if NAPLAN wasn't around", that NAPLAN "Doesn't say more than what teachers already know" (Parent/carers' association), or "Hasn't given us much at all other than an ongoing debate in the country when we should have been talking about equity" (Teachers' union). Others took the view that those wishing to have NAPLAN replaced should be 'careful what they wish for'. As one of the school system/sector stakeholders put it: "If we removed NAPLAN, sectors would need something else to replace the gap. The new testing could be more burdensome on teachers. It's naïve to say 'Let's get rid of it' without considering an alternative."

Some stakeholders argued there have been positive outcomes from NAPLAN. As one representative of a principals' association put it, 'NAPLAN has improved focus on literacy and numeracy on schools – a positive outcome.' Others argued that a decade of NAPLAN testing had had no impact:

It has not contributed to an increase in educational outcomes. It has heaped public scorn on disadvantaged students and communities, which are placed in the modern day stocks through the invasive My School website. It rewards a narrow band of often lower-order intellectual capacities; it has narrowed the taught curriculum; it has corresponded to a seemingly inexorable decline in Australia's performance in major international tests. (Teachers' union)

Whether or not they believed that NAPLAN or the international comparisons had had an impact, most written submissions indicated broad support for some type of national assessment, usually as a tool for helping teachers and schools understand how best to assist students but often also to support trend analysis capable of informing

policy development. As a submission from a statutory authority put it, 'While NAPLAN has a range of issues, as raised in the review's interim report, national standardised testing can serve as important indicator of the health of Australia's school education system'.

Even the harshest critics of NAPLAN acknowledged the importance of having a national standardised testing program:

The teaching profession continues to recognise that it is essential to have a National Assessment program for Australia's students. Such a program is imperative if we are to support communities most in need, to track how educational standards are developing and to assist individual students to grow and progress to their optimal level. (Teachers' union)

Whether or not the current NAPLAN testing program meets these high standards is the topic of subsequent chapters in this review.

# Chapter 3: Other national educational assessment practices

National educational assessment policies and practices vary, and the practices do not relate systematically to the quality of students' learning. This chapter describes practices in countries, selected because they are like Australia in important respects (Canada, New Zealand, England, Scotland) or stand out because they are high-achieving (Finland, Singapore, Japan). It concludes with a description of issues of relevance for consideration in Australia.

## Key points

- There are no assessment practices common across high-performing countries.
- Some have maintained external, subject-based examinations at the end of primary school and in mid-secondary school.
- Some, like Australia, having abandoned external, subject-based examinations before the end of secondary education, have introduced census assessments at particular year levels of foundational skills in literacy and numeracy and, in some cases, of science as well.
- In most cases, school results are provided only to the schools and the education authorities but, in some case, some of the school-level results are made publicly available.
- In most cases, individual students' results go to the students' schools and their families but, in Scotland, they go only to the schools.
- All those without census assessments, and some with them, use sample assessments to monitor the performance of their education system.

## Country assessment policies and practices

### Singapore

Singapore does not conduct assessments of general literacy and numeracy skills. It has retained subject-based, national examinations taken by all students at the end of primary education, in mid-secondary education, and in the final year of secondary education.

The primary school curriculum includes the following subjects: English Language, Mother Tongue Language (MTL), Mathematics, Science, Art, Music, Physical Education, Social Studies, and Character and Citizenship Education (Ministry of Education Singapore, 2020c).

School-based assessments are conducted in all levels of primary education but, from 2019, the extent is being reduced. To explain the grounds for the change, the Ministry of Education (2018) says, 'To meet the challenges of an increasingly complex world, our students need to be lifelong learners. To nurture lifelong learners, we need to help our students discover more joy and develop stronger intrinsic motivation in learning.' Year-end, school-based examinations have been removed from primary (P)1 (Year 1) and P2 (Year 2) since 2019 and mid-year, school-based examinations are being removed from P3 (Year 3), P5 (Year 5), secondary (S)1 (Year 7) and S3 (Year 9) over the period 2019 to 2021. Annual reports to parents/carers in the Holistic Development Profile will no longer provide comparative information on a student's place in class or in relation to a class mean but will still include subject marks and grades, form teacher's comments, ratings of personal qualities and reports on physical fitness, involvement in community-based and co-curricular activities, and school attendance.

At the end of primary education (P6, Year 6), there is an annual national Primary School Leaving Examination conducted by the Singapore Examinations and Assessment Board (SEAB). It involves oral examinations and listening and comprehension examinations in English Language and Mother Tongue, as well as written examinations, mostly between one and two hours, in English Language and Mother Tongue, mathematics and science (SEAB, 2020). The other primary school subjects are not assessed in the national examination. Students are admitted to secondary schools based on merit in the leaving examination and their choice (Ministry of Education Singapore, 2020a).

In mid-secondary education, students sit the Singapore-Cambridge General Certificate of Education examinations depending on their course of study. Students in the Normal (Academic)-Level (GCE N(A)-Level) and the Normal (Technical)-Level (GCE N(T)-Level) sit the examinations in S4 (Year 10). Students in the Ordinary Level (GCE O-Level) sit the examinations in S5 (Year 11), or in S4 if they are taking an Express Course. Students in the Technical Level who excel in specific subjects may be allowed to take the examinations at the GCE N(A)-Level and students in the Academic Level who excel in specific subjects may be allowed to take the examinations at the GCE O-Level (SEAB, 2020). From 2024, the GCE O- and N-level streams will be replaced by subjects grouped into three levels of study; from 2027, the GCE O- and N-Level examinations will be consolidated into a common examination in a Singapore-Cambridge Secondary Education Certificate (Times Online, 2020).

Students' results in GCE O-Level can be used for pre-university entry. They will be admitted to a two- or three-year course leading to the Singapore-Cambridge General Certificate of Education Advanced Level (GCE A-Level) examinations. Students

can choose to be examined at three levels of study – Higher 1, Higher 2 and Higher 3 (Ministry of Education Singapore, 2020b). Admissions to university courses are based on examination performance and additional interviews/tests if required (for example, National University of Singapore, 2020).

## Japan

National assessment of students in Japan was conducted until 1964. It was discontinued during a period of political conflict with the Japan Teachers' Union, which was concerned that national assessment was used by the government to control educational content. In the 1990s, there was considerable discussion of what was taken to be a decline in the quality of student learning attributable to the Ministry of Education, Culture, Science, Sport and Technology (MEXT) policy of *yutori kyōiku* or 'education that gives children room to grow', *yutori* meaning 'relaxed' or 'pressure-free'. The debate was reinforced by a report that university students could not perform calculations with fractions (Okabe, Tose & Nishimura, 1999 quoted in Kuramoto & Koizumi, 2016, p.420). In 2002, MEXT issued its 'Recommendation for Learning' which called for 'an improvement in scholastic achievement' and was seen to be a step back from the relaxed approach (Kōichi, 2012).

The first Program for International Student Assessment (PISA) survey of the achievements of 15-year-olds in 2000, however, suggested that Japanese education was performing well prior to the policy change. Only Finland was significantly better than Japan in reading and none was better in mathematics or science. In PISA 2003, ten countries were significantly better in reading, three in mathematics and none in science. This produced another discussion about declining standards but then improvement or maintenance was achieved in PISA 2006 (nine ahead in reading, four

in mathematics and two in science) and PISA 2009 (five ahead in reading, five in mathematics and three in science). Some of the relative decline was due to other high-performing countries joining PISA – Hong Kong, Taiwan and Estonia from 2006, and Shanghai and Singapore from 2009 (OECD, 2001, pp.53, 79, 88; OECD, 2004, pp.281, 92, 294; OECD, 2007, pp.296, 316, 56; OECD, 2010, pp.54, 134, 151).

In the 1990s, the Curriculum Council was tasked with monitoring the achievement of goals and the content of the Courses of Study. It proposed a comprehensive nationwide survey of academic achievement among a representative stratified sample of students across school years and subjects for each administration. In 2007, the Council on Economic and Fiscal Policy considered the enhancement of academic achievement through competitive principles and proposed the introduction of the National Assessment of Academic Ability (NAAA).

The NAAA was introduced in 2007 for all students but was administered to only a sample of students in 2010 to 2012. It has been administered to all students since then. Students are tested in Grade 6 (end of primary) and Grade 9 in Japanese, mathematics and science, with English added from 2019. Each subject test has a section that assesses comprehension and a section that assesses application skills. Mean NAAA subject scores in each region are announced annually. Average scores are shared with schools and prefectures so that they can identify weak schools or areas of policy that need attention. MEXT, however, requires schools and school boards to publish their improvement plans partly based on data drawn from the national assessments.

Kuramoto & Koizumi (2016) claim there is ambivalence to testing in Japan, largely attributable to the high-stakes university

entrance examinations and the competitive labelling of students based on the university to which they gain admission. This competitiveness is said to influence preparation for the NAAA assessments in the earlier school years with the consequence that “almost every problem with Japanese youth is attributed to testing” (p.418). There are, however, earlier examinations on completion of elementary (Years 1 to 6) and lower secondary school (Years 7 to 9) for entry to upper secondary school (MEXT– Japan, 2020). ‘Admission into senior high schools is extremely competitive, and in addition to entrance examinations, the student’s academic work, behavior and attitude, and record of participation in the community are also taken into account. Senior high schools are ranked in each locality, and Japanese students consider the senior high school where they matriculate to be a determining factor in later success’ (CIEB, 2020).

## Canada – Ontario

Canada is a federation like Australia, but school education is the exclusive responsibility of the provinces and territories. There is no national ministry of education. Some ‘pan-Canadian’ issues are dealt with collaboratively by the provinces and territories through the Council of Ministers of Education, Canada, but assessment is a matter of provincial policy. This description focuses on the most populous province, Ontario. In PISA 2018, Ontario was significantly ahead of Australia in reading, mathematics and science (OECD, 2019b, pp. 73, 76, 79).

At the primary school level, there are ‘curriculum-based, province-wide assessments that measure the reading, writing and maths skills they are expected to have learned by the end of Grade 3 and Grade 6. ... All students who attend publicly funded schools and who follow the Ontario Curriculum are required to write them.’

There are four language sections and two mathematics sections. They measure whether students understand different types of texts; express their thoughts clearly for others to understand; and have acquired the appropriate mathematics skills to solve problems. The assessments are conducted during a three- to six-day period in late May and early June, thus towards the end of the academic year. Students have approximately one hour to complete each section. The assessments are conducted by the Ontario Education Quality and Accountability Office (Ontario EQAO, 2020a), an arms-length organisation to the Ontario Ministry of Education.

Results are available when students return to school after the summer vacation. Students receive an Individual Student Report directly from their school to take home. Parents/ carers and students are assured that there is no need to study for the assessments. ‘EQAO assessments are based on the Ontario Curriculum and do not require additional preparation (for example, tutoring, extra books)’ (Ontario EQAO, 2020a, p. 2).

Student cohorts are tracked from Grade 3 to Grade 6, with results identifying students as ‘maintained standard’, ‘rose to standard’, ‘dropped from standard’ and ‘never met standard’. EQAO emphasises that the assessment results tell only part of the story on students’ learning and progress:

EQAO assessment results should be reviewed alongside students’ daily classroom work and other student-achievement-related assessment information to gauge student learning and determine where more support may be needed. For students who do not meet the provincial standard, it is particularly important for parents or guardians and educators to discuss how to work together to close learning gaps and improve student achievement (Ontario EQAO, 2020a p.3).

At the secondary school level, there is a Grade 9 Assessment of Mathematics which has different versions for students in the academic and the applied mathematics courses. Grade 9 mathematics teachers may use this test as part of their course assessment. There is a Grade 10 Ontario Secondary School Literacy Test (OSSLT) on which successful completion is one of the requirements to earn an Ontario Secondary School Diploma; however, if students are unsuccessful in passing the OSSLT after two attempts they may enrol in a literacy course as an equivalent diploma requirement. There are no provincial subject-based examinations at the end of secondary education.

In September 2017, the Government of Ontario announced a review of provincial assessment and reporting practices. A statement issued in June 2019 indicated that most of the changes involved additional support for students who were English language learners. On the Ontario Secondary School Literacy Test, the report category 'Unsuccessful' would be replaced by 'Not Yet Successful' in student, school, board and provincial reports (Ontario EQAO, 2020b).

In 2012, EQAO introduced EQAO Reporting, an interactive web-based reporting application that enables school principals to access their school's EQAO data and to link achievement data to contextual and attitudinal data. This application was made available to elementary school principals in 2012 and to secondary school principals in 2013.

## England

Traditionally, primary and lower secondary education in England was highly decentralised with control in the hands of school and local education authorities. There were external examinations at the end of 11 years of schooling (for 16-year-olds) for the General Certificate of Education Ordinary

Level (GCE O-Level) in which students sat for examinations in eight to nine subjects. External examinations were also held at the end of secondary education after 13 years of schooling (for 18-year-olds) for the General Certificate of Education Advanced Level (GCE A-Level) in which students studied and sat for examinations in three subjects of considerable depth. The GCE curricula and examinations were conducted by independent examinations boards. Schools determined from which board they took the subjects, including the possibility of taking different subjects from different boards. Over several stages of amalgamation, the board have reduced to three: Assessment and Qualifications Alliance; Pearson; and Oxford, Cambridge and RSA, which all operate under guidelines from the Office of Qualifications and Examinations. The guidelines include specifications of subject content and what the mix of examination and school-based assessment is involved, with the school-based component ranging from 0% (for example, in English, mathematics, chemistry) to 100% in art and design (Ofqual, 2020).

The Education Reform Act 1988 introduced a first national curriculum in England. National assessments were introduced in English, mathematics and science at the end of Key Stage 1 (Years 1 and 2), Key Stage 2 (Years 3 to 6) and Key Stage 3 (Years 7 to 9) when most students are aged 7, 11 and 14 respectively. The GCE O-Level at the end of Key Stage 4 (Years 10 to 11) was replaced from 1987 by the General Certificate of Secondary Education (GCSE) to provide a national qualification for students wanting to leave school at 16 years without going on to GCE A-Level. At Key Stages 1, 2 and 3, schools were statutorily obliged to report on students' performances using standardised assessment tasks (SATs). At Key Stage 1 they were cross-curricular tasks delivered in the classroom while at Key Stages 2 and 3 they were tests.



A new primary curriculum was introduced in 2014. 'End-of-key stage national curriculum tests were re-designed to take account of the national curriculum programmes of study, and to provide more accurate and reliable information for teachers and parents/carers, and for school accountability purposes. ... The new progress measures, introduced in 2016, ensure that schools are recognised for the work they do with all of their pupils, regardless of whether these pupils are high, middle or low attainers' (UK Department of Education, 2017, p. 3).

[Progress measures] provide a much stronger incentive for schools to focus on improving the attainment of the lowest-attaining pupils, rather than focusing efforts on getting pupils over the threshold of the expected standard.

Such progress measures require a baseline to establish pupils' starting points ... to work out how well, on average, a school's year 6 pupils do at key stage 2 compared to other pupils nationally with similar starting points. ... [T]he intention is for a new assessment to be introduced in the reception year to act as this baseline. Roll-out of the assessment on a statutory basis will be in autumn 2020, with a large-scale pilot in the preceding year (p.11).

The new reception measure will be used only to create 'school-level average progress measures when the pupils reach the end of key stage 2, 7 years later'.

With the introduction of statutory reception baseline assessment, assessments at the end of Key Stage 1 will become non-statutory from 2022-2023 and the existing non-statutory English grammar, punctuation and spelling test will remain non-statutory (p.16). A new statutory, national multiplication tables check, however, has been scheduled for introduction in the 2019 to 2020 academic year, with the intention that '[d]

ata from the assessment will be published at national and local level only, not at school level, and data from the check will not be used to trigger intervention or inspection' (p.17).

At the end of Key Stage 2, there is a 'statutory duty for schools to report teacher assessment judgements in English reading, English writing, mathematics and science' but the teacher judgements in reading and mathematics are not used 'to calculate headline accountability measures, as data from national curriculum tests is used instead'. Consequently, and in an effort to reduce teacher workload, the requirement for teachers to assess students 'against teacher assessment frameworks in reading and mathematics' will be removed (p. 14).

## Scotland

Scotland has had regular sample surveys of students' achievements from 1983, initially by the Assessment of Achievement Programme, that mainly assessed English language, mathematics and science (1983 to 2004), then by the Scottish Survey of Achievement (SSA) conducted annually in primary and secondary schools from 2005 until 2009. The SSA collected evidence on students' achievement and progression; teachers' judgements of pupils' attainment levels; students' and teachers' experience of learning and teaching; and changes in performance over time. It focused on a different aspect of the school curriculum each year. In 2011, the SSA was replaced with the Scottish Survey of Literacy and Numeracy (SSLN), which supported assessment approaches under Scotland's new Curriculum for Excellence (CfE). It was conducted until 2016, assessing primary (P)4, P7 and secondary (S)2 learners' progress in literacy and numeracy in alternate years. It also collected information on students' and teachers' attitudes towards aspects of learning and teaching.

The SSA covered literacy and numeracy – but in the context of other subjects – so provided a measure of progress in those subjects as well. It collected information on students' attitudes and also, for example, on practical work in science. Because some local authorities thought the assessments were too time consuming, the survey was restricted to literacy and numeracy. When that survey showed a decline in performance, the local authorities said that it was because it did not give good information. In response, the Scottish Government withdrew the sample survey and developed a census assessment of all students in particular years of schooling.

In 2016, the Scottish Government published 'The National Improvement Framework for Scottish Education'. Based on evidence provided in the development of the framework document, the government nominated six key drivers for improvement, one of which was assessment of children's progress (Scottish Government, 2016, pp. 44-45). The framework was developed to support a new curriculum and intended to 'provide a level of robust, consistent and transparent data across Scotland to extend the understanding of what works, and drive improvements across all parts of the system' (ACER, 2018, p. 6).

The Scottish Government then discontinued the sample-based SSLN and introduced a census collection of Achievement of Curriculum for Education (CfE) Levels as a replacement to inform and target improvement at school, local authority and national level. Teachers of students in P1, P4, P7 and S3 indicate whether each child in their class has achieved the CfE level associated with that stage. Teachers' professional judgement is at the heart of the Scottish Education system (Hayward, 2018, Hutchinson & Young, 2011). However, the government also introduced the

Scottish National Standardised Assessments (SNSA) to provide nationally consistent information about progress in literacy and numeracy as an additional source of diagnostic information to inform teachers' professional judgements.

The Australian Council for Educational Research (ACER) was contracted to develop assessments in numeracy, reading/literacy and writing for use in P1, P4, P7 and S3. Schools administer Scottish National Standardised Assessments (SNSAs) once each year at a time they choose but they may opt out of the program. The assessments are digital and delivered online and 'reports to schools and teachers are provided as soon as a learner completes an assessment. Additional reports are available for local authorities' (ACER, 2018, p. 6). National reports are produced each year noting overall national achievement results and analyses of the achievement levels by gender, ethnic background and for various subgroups of students seen to have special needs (for example, those with additional support needs, registered for free school meals, in out-of-home care, and speaking English as an additional language.) The latest national report, the second to be produced, is for the academic year 2018-19 (ACER, 2020b).

The introduction of the SNSA tests was contentious, particularly with young children. In a review of testing at P1, Reedy (undated), however, concluded that, while 'media reports and some members of the Scottish Parliament reported that the P1 SNSA was causing children distress ... the majority of head teachers and teachers did not see any distress or discomfort as children undertook the P1 SNSA, in fact, they reported that the children enjoyed it' (p. 39).

The Scottish assessments are similar to NAPLAN in their coverage of reading, language conventions, writing and numeracy and in their application to all students in a census in four years of schooling. The Scottish assessments cover P1, 4, S7 and S10 whereas NAPLAN assesses in Years 3, 5, 7 and 9. Like NAPLAN Online, the Scottish assessments are computer-delivered as adaptive (branching) tests. Individual student reports are similar to the NAPLAN student reports. There is a continuum drawn up the page with bands on it and descriptions of what a student at a particular band can do. The students' location on the band is marked but there are no markers for school, local authority or national means. There are three other significant differences. One is that Scottish schools can decide whether to administer the tests. The national reports on the program do not indicate whether any schools did opt out, but they report that 95% of students were tested in 2017 to 2018 (ACER, 2018, p.12) and 93.4% were tested in 2018 to 2019 (ACER, 2020b, p.12). These participation rates match those achieved with the NAPLAN census tests in Australia. Schools can also decide when to administer the tests, although some of the 32 local authorities insist that all students take the tests at the same time. This variation in time of testing means that national means would not be useful in any case. The other, more substantial, difference is that the results are not published and are not provided to parents/carers and students unless the school chooses to do so. They go only to the school and the teacher to become one piece of information that teachers use with their own local information for their assessments of students and reports to parents/carers. The information collected centrally from schools is teachers' judgements of their students based on their own assessments and the students' SNSA results.

## New Zealand

New Zealand has no nationally mandated external assessment of all students until the end of secondary schooling, though the National Administration Guidelines state that 'Each board of trustees, with the principal and teaching staff, is required to: ... on the basis of good quality assessment information, report to students and their parents/carers on progress and achievement of individual students in plain language, in writing, and at least twice a year and across the National Curriculum ... including in mathematics and literacy' (NZ Ministry of Education, 2020c). The Ministry provides detailed information on assessment tools and resources (NZ Ministry of Education, 2020a).

From 1997, the Ministry of Education provided a School Entry Assessment that school could use to assess students' concepts about print, numeracy and their oral language. A replacement assessment is being developed.

A sample-based National Education Monitoring Project operated from 1995 to 2010. It was replaced in 2012 by the National Monitoring Study of Student Achievement, which assesses students in Years 4 and 8 in arts, health and physical education, science, English, mathematics and statistics, social sciences, technology, and languages, with the subjects rotated over a five-year cycle. The assessments involve a mix of group and individually administered tasks, with some administered on computers. The program is a collaboration between the Educational Assessment Research Unit at the University of Otago, the New Zealand Council for Educational Research (NZCER) and the Ministry of Education (University of Otago, 2020).

As an alternative to external testing, new National Standards were introduced in 2010. The justification was that there was 'an urgent need to raise student achievement and for parents/carers to be better informed

about their children's performance in literacy (reading and writing) and numeracy in their primary and intermediate schooling years'. Schools were required to use the standards to guide teaching and learning, to report children's progress and achievements against the standards to parents/carers, and to include baseline data and targets in their 2011 Charters. Annual reporting of results was required from 2012. Some of these assessments are available in Māori, together with others exclusively in Māori (NZ Ministry of Education, 2020a).

A Progress and Consistency Tool (PaCT) was introduced in 2015 to support teachers in making dependable judgements about their students' achievement. It provides decision frameworks that capture teachers' 'best fit' judgements of their students on aspects of mathematics, reading and writing. It locates the students' overall level of achievement on scales on which progress can be tracked from school-entry to Year 10. The PaCT scales were originally benchmarked against National Standards but are now linked to curriculum levels (Education Services, NZ, 2020).

Judgements against the standards were taken to be comparable across schools and they were used, among other things, to create league tables. They were also claimed to narrow the curriculum and, after a change of government in 2017, were abolished in favour of 'plain English' reporting to parents/carers on students' progress without reference to National Standards (Collins, 2017).

Teachers and schools use a range of externally developed assessments (NZ Ministry of Education 2020a). These include NZCER's Progressive Achievement Tests in reading, listening comprehension, punctuation and grammar and mathematics (NZCER, 2020b) and the Assessment Tools for Teaching and Learning (e-asTTle), which assess reading, mathematics and writing (NZ Ministry of Education 2020b). With e-asTTle, teachers

can design their own tests by assembling groups of items from an item bank. Multiple-choice questions are machine-marked online. Open-ended questions are marked by the teacher against a marking guide and the results entered online. Schools can compare their results with the curriculum levels. The Ministry has also developed Assessment Resource Banks in mathematics, English and science, administered by NZCER (2020a). These materials have a strong formative focus.

## Finland

In Finland, there are regular sample-based surveys with standardised tests of student learning outcomes in pre-primary (one year) and basic education grades 1 to 9 (age 7 to 16). The surveys are conducted by the Finnish Education Evaluation Centre (FINEEC). There were, for example, surveys of English in grade 7 in 2018 and mathematics at the end of basic education in 2020. A survey of English in grade 9 is scheduled for 2021. The samples involve 5% to 10% of the relevant age group, with oversampling of schools providing education in Swedish, the second national language, to obtain stable estimates for that sub-population.

The assessments are based on objectives defined in the national curriculum. The evaluation tasks are trialled in schools that are not in the sample and the final evaluation instrument is put together based on feedback from teachers and analyses of the data from the trial. Schools in the sample receive information on their students' performances in relation to the national average. Information is also collected from principals, teachers and students 'on working methods and teaching arrangements, educational resources, student evaluation and study attitudes of the pupils' (FINEEC, 2020). The responses from the students feed into an established indicator to study students' views of themselves as learners of the subject,

on the attractiveness of the subject and on the usefulness of studying it. A national report is prepared and summaries are provided to meet 'the needs of the Ministry of Education and Culture, the Finnish National Agency for Education, Departments of Teacher Education, education providers, schools, teachers and other bodies' (FINEEC, 2020).

Apart from those in sample schools, students do not take any external assessment until the end of secondary education and then only if they wish to go on to university. The focus of assessment in the schools is formative to facilitate and guide students' learning. At the end of general upper secondary education (grades 10 to 12), which is the path to university education, there is a Matriculation Examination in which students take four examinations – mother tongue, and three by choice from second national language, a foreign language, mathematics and one from humanities and natural sciences. At the end of vocational upper secondary education, 'Qualification-specific learning outcomes evaluations focus at vocational skills and are based on vocational skills demonstrations and supplementary evaluation material, such as students' self-evaluations, self-evaluations of VET providers and workplaces, and evaluations of the quality of the demonstrations' (Matriculation Examination Board, Finland, 2020).

FINEEC conducts thematic and system evaluations that focus on 'the state of a certain form of education ... a whole education system or some part of it ... education policy and its implementation

or the renewal and development processes of the education system. Evaluations may target one educational level or cover several'. Recent evaluations have included 'learning and competencies in basic education and general upper secondary education', 'best practices for the integration of immigrants into the educational system' and 'student transitions and smooth study paths' (FINEEC, 2020).

## Potential relevance for Australia

Table 10 compares the mean performances of Australian students with the mean performances of students in the countries described in the previous section on the three main PISA scales – reading, mathematics and science in 2018. Canada–Ontario, Finland and Singapore were significantly ahead of Australia in all domains; Japan was significantly ahead in mathematics and science but not different in reading; England was significantly ahead in mathematics but not different in reading and science; New Zealand was significantly ahead in science but not different in reading and mathematics; and Scotland was not different from Australia in reading and mathematics but significantly behind Australia in science (OECD, 2019b, pp. 57, 59, 61, 73, 76-77 & 79).

The countries for which assessment policies and practices have been summarised in the preceding section are higher performing than Australia or equal to Australia except for Scotland in science where its performance is significantly lower than Australia's.

**Table 10:** Position of comparison countries in relation to Australia in PISA 2018

Location	Reading	Mathematics	Science
Canada - Ontario	▼	▼	▼
England	•	▼	•
Finland	▼	▼	▼
Japan	•	▼	▼
New Zealand	•	•	▼
Scotland	•	•	▲
Singapore	▼	▼	▼

**Key:** • Not different from Australia ▲ Significantly behind Australia ▼ Significantly ahead of Australia

The interesting question, from the perspective of this review of NAPLAN, is whether there are common features of their assessment policies that are significantly different from Australia's, which might account for their higher performances. Australia's policies provide for census assessments of students in NAPLAN in Years 3, 5, 7 and 9 in reading, language conventions, writing and numeracy, as well as assessments of samples of students on a three-yearly cycle in Years 6 and 10 in science, civics and citizenship, and information and communication technology literacy. NAPLAN results are reported publicly at national and state and territory level and for subpopulations of interest, including for males and females, Indigenous students and students with a language background other than English. Reports on students are provided to schools, parents/carers and students and reports on schools are provided to education systems/sectors and schools and were, in earlier versions of My School, made available publicly in comparison with the results of other schools with students from a similar level of socio-educational advantage. A comparison with the other countries is provided in Table 11.

This table, together with the descriptions in the summaries of country policies make it clear that there are no assessment policies or practices common to all the countries surveyed. All of them have assessments

of all students at the end of secondary education in whatever subjects they are studying. Australia does too but that is outside the range of NAPLAN testing. The comparisons offered in the text of this chapter and in Table 11 are for the years prior to the end of secondary school.

There are high-performing countries in Australia's region, Singapore and Japan that retain subject-based examinations at the end of primary education and in mid-secondary education. Ontario is a provincial system like one of Australia's large states. It conducts census assessments quite like NAPLAN and reports all results down to school-level publicly. That is what Australia did with NAPLAN results prior to the recent changes to the My School website, which no longer gives such visibility to school comparisons. The website still provides the school means, with confidence intervals, on the results page and the comparison of school means to means of students with a similar background visible with a hover-over function. The dominant reports on the My School website for individual schools are graphical representations of growth rates for students between Years 3 and 5 and between Years 7 and 9. England has census assessments and now, like Australia, focuses on growth not current status. It provides school-level information to the schools, as well as to central and local authorities, but not publicly.

Scotland offers cohort assessments in numeracy, reading and writing but school participation is optional. Students' results are reported to schools, but they are not provided to students or parents/carers. It is expected that teachers will use the students' results from the tests together with their own information on each student for reports to parents/carers and as part of school reports to local authorities.

Finland and New Zealand have sample assessments that are conducted in all school subject domains on a cycle. Finland adds a cycle of school reviews. New Zealand provides schools and teachers access to a range of assessment resources for local use. Finland is a high-performing country that has no census assessments of students. It has a cycle of sample assessments that covers all school subjects and thematic and system evaluations.

Finland, New Zealand and Scotland are rather similar in their current policies and practices, but Finland has attracted more

international attention because it has much higher results in PISA, and because its current policies and practices are more longstanding. Finland and New Zealand have only sample assessments at a national level. Scotland is attracting some interest now because it does have census assessment, despite an 'opt-out' provision for schools, to obtain measures on all students but does not provide the results beyond the school to students and parents/carers. Instead, teachers use the results from the Scottish National Standardised Assessments as one piece of evidence alongside the teachers' own assessments.

Finland gained a lot international attention when its students performed very well in the first PISA in 2000. It was significantly ahead of all others in reading, significantly behind only Japan in mathematics and significantly behind only Korea in science (OECD, 2001, pp.53, 79, 88). Over the successive PISA cycles since 2000, however, Finland's performance has declined significantly in all domains (OECD, 2019b, pp. 57, 59, 61, 131).

**Table 11:** Nature of assessments in other countries

	Finland	Japan	Singapore	Ontario	New Zealand	England	Scotland
<b>Test population</b>	Sample	Census	Census	Census	Sample	Census	On demand
<b>Test domains</b>	All school subjects over time	Japanese Maths Science English	English Language, Mother Tongue, Maths, Science	Reading Writing Maths	All school subjects over a five-year cycle	English Maths Science	Numeracy Reading Writing
<b>School years assessed (Australian equivalent)</b>	Years 1-9	Years 6, 9	Years 6, 10	Years 3, 6, 9, 10	Years 4, 8	Years 2, 6	Years 1, 4, 7, 10
<b>System results public</b>	✓	✓	✓	✓	✓	✓	✓
<b>Students' results to parents/carers</b>	✗	✓	✓	✓	✗	✓	✗
<b>School-level test results to school</b>	✗	✓	✓	✓	✗	✓	✗
<b>School-level test results to public</b>	✗	✗	✓	✓	✗	✓	✗

Many educators from other countries, including many Australians, have visited Finland since the PISA 2000 results were published in 2001, in the hope of learning lessons for their own countries. Frequently, they cherry picked those policies and practices that coincided with their personal preferences for their own domestic policies. Among them, the selectivity of teacher education programs, high quality of teachers, a high-level curriculum leaving a great deal of discretion to schools and teachers and the absence of external assessments of students until the end of upper secondary education for the students in the general education stream. Visitors presumed that they were seeing in Finland's current policies the seeds of its success in 2000. They may have been seeing the seeds of its decline to 2018.

Sahlberg, former Director-General of the Centre for International Mobility in the Finnish Ministry of Education and Culture, warns against simple notions of transfer between systems (2015, p.xxii) and emphasises the need to understand the complex series of reforms over time that stood behind Finland's high performance in PISA in 2000 (p.5). Oates (2015) points out

that, while the Finnish national curriculum is a general, high-level document, government-approved textbooks gave detail until 1992 and continued to be used after that. Reflecting on developments in Finland and in other countries, Sahlberg (2015, 2016) has provided a list of features of reform in other countries that he claims are different from Finland's and counterproductive – 1) competition among schools for enrolment, – 2) standardisation of teaching and learning, – 3) increased emphasis on reading literacy, mathematics and science, – 4) borrowing change models from the corporate world, – and 5) test-based accountability policies. He invites those learning from Finland to reject these assessment practices. They are practices that characterise Japan, Singapore, Ontario and England.

The more general conclusion of the survey of countries covered in this chapter is that there are no common assessment practices in high-performing countries. In the end, each will need to develop its own policies and practices while, examining the practices of others.



# Chapter 4: Quality of NAPLAN digital tests

NAPLAN assesses all students in Years 3, 5, 7 and 9 in reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. From 2008 to 2017, all the tests were delivered on paper. From 2018, a growing proportion of schools administered the tests online in digital format, which were marked by the computer as the students took the tests. The exception was writing, which was marked by human makers. This chapter considers the properties of the digital tests, to some extent in comparison with their paper precursors. The writing tests are considered in Chapter 5.

## Key points

- From 2008 to 2017 NAPLAN tests in reading, language conventions (spelling, grammar and punctuation) and numeracy have involved both multiple-choice questions (for which answers can be machine scored) and short, constructed-response formats that humans mark.
- From 2018, a growing number of schools have used a new computer-delivered digital version of the tests. In NAPLAN Online, all items are multiple-choice and students' responses are recorded by the computer as right or wrong as students respond. The current expectation is that all schools will use this format from 2022.
- From 2008 to 2016, the NAPLAN tests were based on the national Statements of Learning for English and Statements of Learning for Mathematics. From 2017, they have been based on the Australian Curriculum.
- NAPLAN Online delivers branching tests in which, on the basis of the computer's scoring of students' answers as they work through a test, branches the students to items of different complexity after one third of the items have been answered and again after two thirds have been answered. This 'adaptive' testing provides better measures of high and low achievers than can be obtained when all students answer the same questions.
- The use of some common items in the Years 3 and 5 tests, the Years 5 and 7 tests and the Years 7 and 9 tests enable all students' results, in a process called 'vertical scaling', to be expressed on a common scale. Links back to the first NAPLAN enable the results to be expressed on the scale that was originally established in 2008 in a process called 'horizontal scaling'.

## Key points continued

- There is uncertainty in all educational measurements due to uncertainty in the measure itself and, in the case of NAPLAN, to uncertainty in the vertical and horizontal scaling. The level of uncertainty depends on how much data the measures are based on and on how far from the overall mean they are. The most precise means are national, followed by state and territory means. The least precise are individual student's results. Means for large schools are more precise than means for small schools. Means for schools and results for students closer to the national mean are more precise than for those further from it.
- NAPLAN tests are currently census tests intended to be taken by the whole-cohorts of students in Years 3, 5, 7 and 9, with some specific adjustments to accommodate students with disabilities. There are provisions for parents/carers to request students be withdrawn from the testing and there are students absent for other reasons on the days of testing. Non-participation rates vary across the states and territories and year levels. Where the rates are high, that may bias estimates of state and territory means.

## Content of tests

### Paper tests

The NAPLAN tests in reading, language conventions and numeracy were paper tests for all students from 2008 to 2017 and, in 2018 and 2019, for students in schools not opting to use the new digital tests provided as NAPLAN Online. The paper tests consisted of multiple-choice items and constructed response items that required a numeric answer, a word or a short phrase. Responses to the multiple-choice items were recorded on a machine-readable form and were machine marked. Responses to the constructed response items were marked by human markers trained to apply nationally agreed marking protocols and item-specific answer criteria (ACARA, 2020b). The structure and coverage of the paper tests in 2019 are shown in Table 12 (ACARA, 2020e, pp. 27-28).

Information on the nature of the tests is provided in Chapter 2. Copies of past NAPLAN test papers and answers for all years from 2008 to 2016 are provided online for teachers, students, parents/carers and anyone else interested in reviewing them. Later tests are not provided because the Australian Curriculum, Assessment and Reporting Authority (ACARA) keeps the specific questions secure 'for a range of purposes, including ACARA's research and development studies' (ACARA, 2020b). Keeping the items secure also enables items to be reused to establish links between the results in different years and to enable the results for successive years to be located on the same scale.

**Table 12:** Structure of paper NAPLAN paper tests, 2019

		Number of items	Time available
<b>Year 3</b>			
Reading		37	45 minutes
Language conventions	Spelling	25	50
	Grammar and punctuation	25	
Numeracy (no calculator)		36	45 minutes
<b>Year 5</b>			
Reading		39	50 minutes
Language conventions	Spelling	25	50
	Grammar and punctuation	25	
Numeracy (no calculator)		42	50 minutes
<b>Year 7</b>			
Reading		50	65 minutes
Language conventions	Spelling	25	50
	Grammar and punctuation	25	
Numeracy	No calculator	40	48
	Calculator allowed	8	
			55 minutes 10 minutes
<b>Year 9</b>			
Reading		50	65 minutes
Language conventions	Spelling	25	50
	Grammar and punctuation	25	
Numeracy	No calculator	40	48
	Calculator allowed	8	
			55 minutes 10 minutes

Some stakeholders regretted the loss of access to the actual tests because they also received the actual responses of their students to the individual items and could see precisely where errors were made. They would then use that information to provide additional instruction to individual students or to adjust their teaching to the whole class. As one stakeholder said:

Previously, teachers could see items, conduct item analysis and have conversations and learn from the tests. Now there is a high-level item

description, link to the Australian Curriculum and a link to an example item. (Education expert)

There was an alternate view expressed from the context of an overseas assessment system.

Never allow information at the item level to be reported. Instead, cluster the items into curriculum concepts and the report [to schools and teachers] can then focus on curricula and not items. (Education expert)

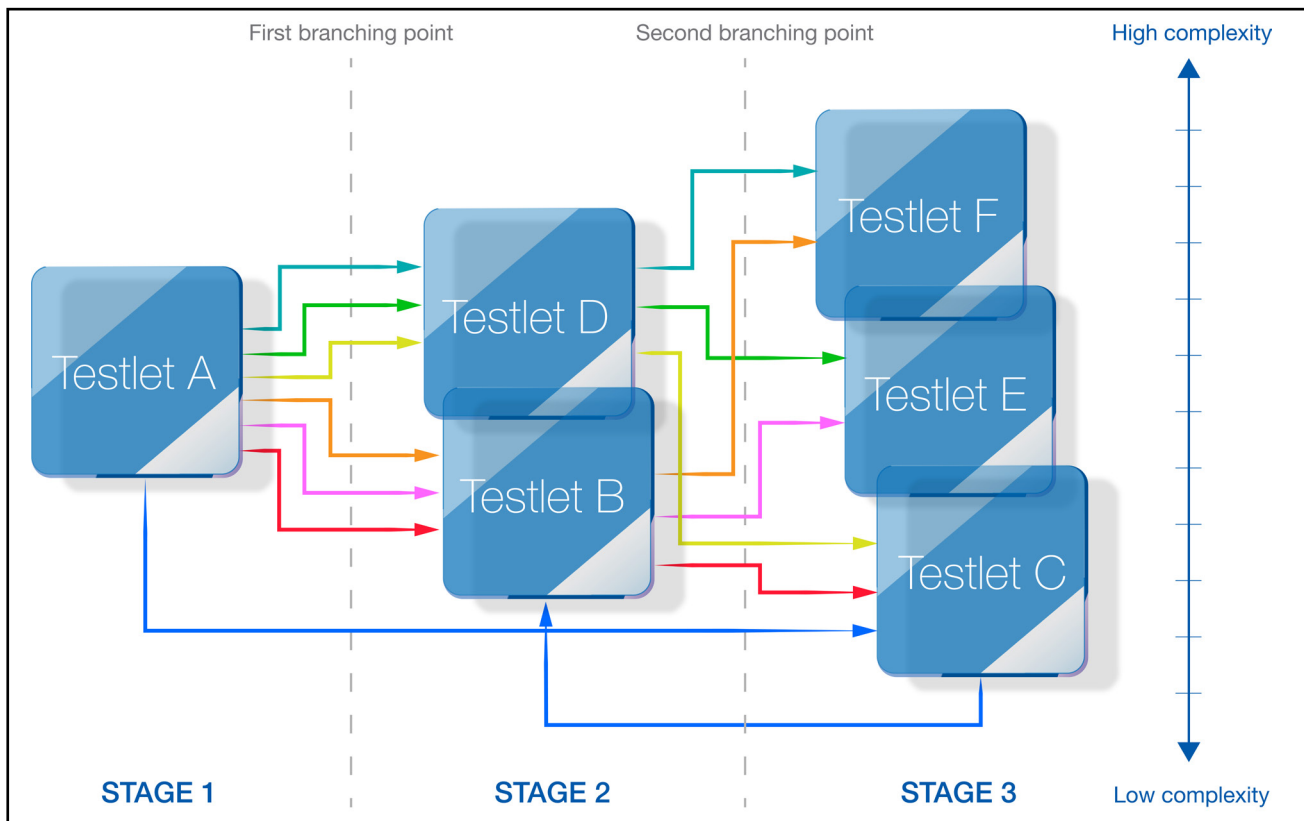
## Online tests

From 2018, there has been a phased shift from paper to online NAPLAN tests. In 2018, just over 15% of schools participated in NAPLAN Online. In 2019, more than 50% did. With NAPLAN cancelled in 2020, the target date for all schools to undertake NAPLAN Online is now 2022, not 2021 as earlier envisaged (ACARA, 2020h).

With the paper NAPLAN tests, all students take the same test. For high-performing students, easy items provide essentially no

information on how well they can perform. Similarly, for low-performing students, difficult items on the common test tell little about how well they can perform.

A key advantage of an online test is that it can be adaptive and present students with tasks targeted close to their performance level and so provide much better information on what each student knows and is able to do. The branching structure for the NAPLAN Online literacy and numeracy tests is shown in Figure 2 (ACARA, 2020e, p. 30).



**Figure 2:** Branching structure of NAPLAN Online literacy and numeracy tests

All students start with common items for their year level in a first testlet A. As they respond to the items, the computer scores each response 'right' or 'wrong'. Students determined by the computer to be doing well at the end of testlet A are moved to testlet D with more complex items while those doing less well are moved to testlet B with less complex items. Students who struggle with testlet A are moved directly to testlet C to give them an opportunity to achieve success on the least complex items before then moving to testlet B.

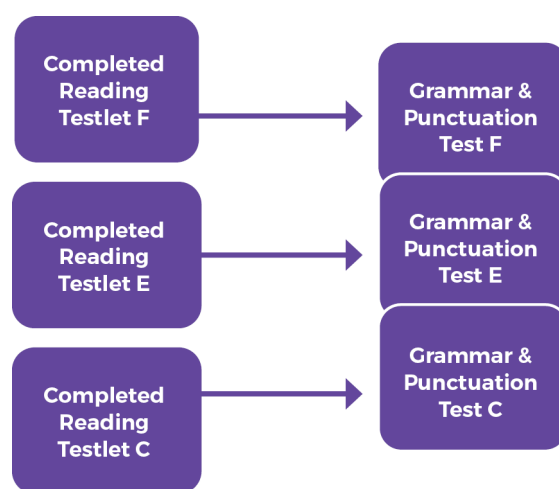
After completing testlets D or B, students are moved to a third testlet based on their performance in their first two: AD or AB. The students performing at the highest level are moved to testlet F, which has high-complexity items. Those performing at the lowest level are moved to testlet C with easy, low-complexity items. Those in between are moved to testlet E. Those whose first pair was AC are moved, as indicated, to B.

There were three versions of each testlet with the versions being comparable in the difficulties of the items, curriculum coverage and skills assessed. The first version of each testlet included items from the paper test and new online items. The other versions included items from NAPLAN 2018 to enable the results to be placed on the same scale as that used in 2018 and ultimately back to 2008, to which it had been linked.

The spelling test had a structure similar to that in Figure 2, except that there were only two testlets at the Stage 3. The testlets in Stages 1 and 2 involved spelling spoken words delivered by the computer. The testlets in Stage 3 involved proof reading to detect spelling errors in text.

The grammar and punctuation test had no branching. Instead it had three tests of differing complexity to which students were directed based on their final testlet in the reading test. The structure is shown

in Figure 3. Students who had completed reading testlet F were directed to a high-complexity grammar and punctuation test F, those who had completed reading testlet E were directed to a medium-complexity grammar and punctuation test E, and those who had completed reading testlet C were directed to a low-complexity grammar and punctuation test C. To link the results onto a common grammar and punctuation scale, there were common items in the grammar and punctuation tests F and E and in the grammar and punctuation tests E and C.



**Figure 3:** Branching structure of NAPLAN Online grammar and punctuation test

Reactions to the online format, from stakeholders contributing to this review, were generally positive – ‘Supportive of the branched testing, even at the item level, though acknowledged this kind of test is difficult to create’ (Subject association); and ‘Appreciate the branching in online testing’ (Member of the NAPLAN Review Practitioners’ Reference Group).

NAPLAN, especially with the online branching model, is designed to direct high-achieving students to their limit by increasing item complexity until students fail. It is contended that this assessment-methodology difference is not well understood and may indeed contribute to the negative feelings reported around NAPLAN testing. That said, for low-

mid achieving students, schools have anecdotally reported more positive test experiences due to the platform design being less confronting than a large paper test, and that the branching method gives opportunity for all students to demonstrate proficiencies in matching with their abilities. (Written submission response: school system/sector)

There were anecdotes about Year 9 students, who had found the NAPLAN tests in Years, 3, 5 and 7 relatively easy and had finished them quickly, being surprised at how much longer they were taking in Year 9 with the online version. Though they had been told about the new form, they apparently did not appreciate that they were slow because they were taking a more complex test.

There was concern about the impact of NAPLAN becoming exclusively online on students in schools with poor connectivity to the internet and for students who do not have access to computers at home to develop fluency in using them.

There will be some students who potentially will never be able to do the test online due to connectivity issues (rural/remote). If the test could be electronic not online (e.g. USB) that may be fine. (Parents'/carers' association)

For these students, the obvious solution to poor connectivity at school at the time of testing would be to provide the tests on a portable device that can be accessed locally at the school. Limitation in access to computers outside school and opportunities to develop fluent use is a serious problem but of a different order and not relevant only for NAPLAN.

One respondent speculated about the possibility of having a paper version of the branching test. That would not be possible because the branching depends on being able to mark students' responses as they work through a test.

Prior to the adoption of NAPLAN as a common national test, the Northern Territory used differentiated tests as an approximation of what branching can provide. Two versions of the paper test were provided with one having more difficult items and the other easier ones but with the two versions having sufficient common items for results to be expressed on the same scale. Teachers were asked to give each student what they judged to be the most appropriate test. A similar model was advocated by the Northern Territory when the form of the new common NAPLAN tests was being negotiated in 2007 but the Northern Territory could not persuade the other jurisdictions to adopt it. With the transition from paper to branching digital tests, the Northern Territory is gaining even more than the flexibility it had lost from 2008.

Since students respond to different items, depending on which path they follow through NAPLAN Online, comparisons cannot be made among students based on the proportion of items they answer correctly. The meaningful measure is a student's score on the underlying NAPLAN scale on which items are arranged by difficulty and students are arranged by achievement level.

The interpretation/analysis of online test items has become more difficult for teachers due to the online test. The paper test showed teachers the proportion of items that were right and they could compare to similar schools and gauge class performance. In the online test you cannot compare a proportion of items [correct] due to branching. (Education expert)

While NAPLAN is conducted as both paper-based and online assessments, every effort is made to make the tests parallel. One consequence is that the power of online assessment cannot be fully exploited, particularly the facility to use more interactive items. The one exception in the transition period is in the spelling tests where spelling of words presented by dictation has been added in the online form to proofreading for spelling errors that has been the means of testing spelling in the paper test.

The structure and coverage of the online tests in 2019, shown in Table 13, are essentially the same as the structure and coverage of the paper tests in 2019 shown in Table 12 (page 59).

## **Item selection**

All items are trialled before they are included in the final tests. Careful analyses are undertaken to detect any bias in items that would disadvantage males or females, students from a language background other than English, Aboriginal and Torres Strait Islander students and students from different states and territories. Bias cannot be judged based on whether items might be easier for some groups than others because that may reflect real differences in achievement levels. Bias is detected through examining relative difficulties of items for the different groups.

**Table 13:** Structure of NAPLAN Online tests, 2019

			Number of items	Time available
<b>Year 3</b>				
Reading			39	45 minutes
Language conventions	Spelling: audio dictation	25	50	45 minutes
	Spelling: proofreading			
	Grammar and punctuation	25		
Numeracy (no calculator)			36	45 minutes
<b>Year 5</b>				
Reading			39	50 minutes
Language conventions	Spelling: audio dictation	25	50	45 minutes
	Spelling: proofreading			
	Grammar and punctuation	25		
Numeracy (no calculator)			42	50 minutes
<b>Year 7</b>				
Reading			48	65 minutes
Language conventions	Spelling: audio dictation	25	50	45 minutes
	Spelling: proofreading			
	Grammar and punctuation	25		
Numeracy	No calculator	40	48	65 minutes
	Calculator allowed	8		
<b>Year 9</b>				
Reading			48	65 minutes
Language conventions	Spelling: audio dictation	25	50	45 minutes
	Spelling: proofreading			
	Grammar and punctuation	25		
Numeracy	No calculator	40	48	65 minutes
	Calculator allowed	8		



For example, in exploring the possibility of gender bias for a particular test, such as Year 7 numeracy, the question is not whether females find the items harder or easier than males. The question is whether the relative difficulty of individual items compared with other items is the same for males and females. If an item stands out in providing a view of male/female differences in performance that is inconsistent with the view provided by the other items, then the inconsistent item would be judged to be gender biased and so excluded from consideration for inclusion in the final test.

Items that are inconsistent with other items are detected using Differential Item Functioning (DIF) analyses. The procedure is discussed and results from its application are provided in the annual NAPLAN technical reports (for example, ACARA, 2020e, pp. 94-100). The DIF analyses behind the development of the NAPLAN tests work effectively to deliver unbiased tests.

## Links to the Australian Curriculum

As described in Chapter 1, all of the states and territories conducted census testing of students in literacy and numeracy prior to the introduction in 2008 of NAPLAN as a common national assessment. When the 2008 NAPLAN tests were developed in 2007, they were based on the Statements of Learning for English (Curriculum Corporation, 2005) and the Statements of Learning for Mathematics (Curriculum Corporation, 2006).

“Since 2016, NAPLAN tests have been aligned to the Australian Curriculum: English and the Australian Curriculum: Mathematics” (ACARA, 2020g). Detailed mapping of the paper tests and the Online Tests in 2019, including mapping by pathway for the online tests, are provided in the NAPLAN 2019 Technical Report (ACARA, 2020e, pp. 34-41). One stakeholder noted, somewhat ironically, the importance of the

link between NAPLAN and the Australian Curriculum, – ‘What are we testing if it is not linked to the Australian Curriculum?’ (Member of the NAPLAN Review Practitioners’ Reference Group)

Other stakeholders, however, revealed the link not to be widely understood. There were comments in the submissions and the consultations that suggested that such alignment would be a good idea. These included, ‘NAPLAN is missing links to the Australian curriculum.’ (Teachers’ association) and ‘Linking NAPLAN to the Australian Curriculum could be a positive development. This may help increase confidence in the test.’ (School system/sector)

There were suggestions, however, that the link was not well-established for the language conventions test.

The grammar in NAPLAN does not match the Australian Curriculum and there are many items in NAPLAN that would only be recognised by children trained in a particular form of language conventions. (Educational organisation)

Some thought the alignment could be to more than the Australian Curriculum: English and the Australian Curriculum: Mathematics. Comments included, ‘It would be a positive move to recalibrate the NAPLAN tests back to the Australian Curriculum. We could assess literacy as part of science.’ (School system/sector); and ‘NAPLAN should be totally aligned to the Australian Curriculum – including general capabilities.’ (Written submission response: principals’ association)

There should also be stronger alignment with the assessment framework and the Australian Curriculum. A continuous improvement model for standardised testing in Australia would consider the benefits of a broader assessment

including General Capabilities, realising a faster turnaround of data and more engaging targeted assessment to deliver a richer, more holistic results dataset more reflective of the Australian Curriculum and better-informing student learning growth strategies. (Written submission response: school system/sector)

Others also picked up on the final part of the preceding comment, expressing a view that strong links to the Australian Curriculum could help teachers identify weaknesses in their own teaching – ‘It should be used to assist teachers to reflect on their teaching of the Australian Curriculum.’ (Written submission response: principals’ association)

There was also a suggestion that communication with parents/carers about NAPLAN results would be more helpful if NAPLAN’s links to the Australian Curriculum and pedagogy were made clearer.

Schools could assist parents/carers and students in their understanding of NAPLAN by better disseminating information about NAPLAN; its place in the curriculum, the process and how the results are used to inform pedagogy and policy. (Written submission response: Education expert)

## Psychometric properties of the tests

### Effect of branching within online tests

There are three main reasons for moving to online testing. One is to achieve rapid scoring and reporting of results to students and schools. A second is to capitalise on the flexibility of electronic delivery to create more complex and interesting test items. The third is to make the tests ‘adaptive’ with students being presented with items close to their achievement level to maximise information about their level.

The adaptive, branching structure in NAPLAN Online provides a further benefit beyond better measurement of students at the extremes of the distribution who are not well served by a common test taken by all students. Having all students taking items that are better matched to their achievement level yields more precise measurement of students throughout the range of performances.



**Figure 4:** Proportions of students taking each path – Year 3 numeracy, 2019

The proportions of students being directed by their performances to each of the pathways in the Year 3 numeracy test are shown in Figure 4 (ACARA, 2020e, Appendix A.1). More than half (57.8%) the students were directed from testlet A to testlet D. From there, half (28.9%) went to testlet F and half (28.9%) to testlet E. The fact that none needed to be directed to the easy items in testlet C suggests that the directions from A to D were appropriate.

Just under half (41.9%) were directed from testlet A to testlet B (27.8%) or direct to testlet C and then back to testlet B (13.8%). Of those who went to testlet B, most then moved on to testlet E (23.8%) and most of the rest moved on to testlet C (4%). The fact that so few went to testlet F suggests that the directions from A to B were appropriate.

The proportions of students being directed to the various paths are determined by the branching rules adopted in the program. A 50:50 split between D and B from A would be desirable. If testlet A works appropriately,

the paths ADC and ABF should be unlikely to be followed, as was the case.

There are seven paths through the branching test structure as shown in Figure 2 and summarised in Figure 4. There are three versions of each of the testlets so students following the same one of the seven paths are likely to be answering different questions. There are, in fact, 126 different paths.

The effectiveness of the branching in achieving better measurement in the extremes of the range of student performances can be seen in Figure 5, where the results for the online assessment in Year 3 numeracy in 2019 are shown (ACARA, 2020e, Appendix A.2).

The first testlet gives an approximation of the distribution of achievement levels that would be produced if all students took the same test as they do with the paper test. That distribution is labelled AXX in Figure 5. By the end of the second testlet, those moved to D (ADX in the figure) have

been differentiated from those moved to B (ABX in the figure). The more difficult items in D reveal the higher achievements of those students. The reason that the ABX distribution does not extend as far down as the AXX distribution is that 13.9% of the students in AXX were directed to testlet C and not to either B or D.

By the time all students have completed their third testlet, the whole population has become much better differentiated than could be achieved with a single, common test taken by all. The subgroup taking the ADF path, for example, extends into a region that would not be measured well by a common test without differentiation. The same can be said of the subgroup taking the ACB path. The branching clearly achieves the purpose of measuring over a fuller range of student achievements.

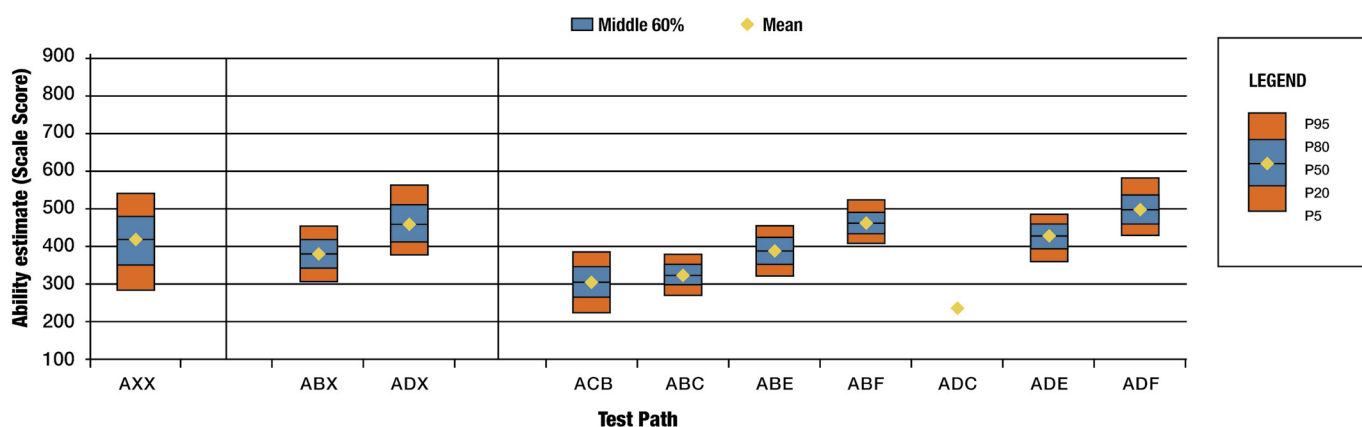


Figure 5: Distributions of student achievement by pathway – Year 3 numeracy, 2019

## Scaling of results over year levels and time

In the first NAPLAN in 2008, students' results across Years 3, 5, 7 and 9 were set on a scale with an overall mean of 500 and a standard deviation of 100. All results were located on the same scale using some common items in the tests for adjacent Years: 3/5, 5/7 and 7/9. In NAPLAN, this common-item scaling is called vertical equating.

Results in subsequent years are also located on the NAPLAN 2008 scale in a process called horizontal equating. Because all NAPLAN items have been publicly released up until 2016, it has not been possible to include items from earlier tests in a current test to use them in common-item scaling for the horizontal equating. Instead, common-person scaling has been used. For each year level, there are secure NAPLAN tests for each domain, developed in 2009, that are administered to samples of students at each year level in the current year who also take the full current NAPLAN tests. Common-person scaling involves the use of the scores of the samples of students in the secure tests from 2009 and the current tests. The process is described for 2017, the last year in which all students did the NAPLAN tests on paper, in the NAPLAN 2017 Technical Report (ACARA, 2018e, pp. 39-49).

The common-person scaling used in the horizontal equating also provides information for vertical equating for the current year because there are common items in the secure 2008 tests for 3/5, 5/7 and 7/9 taken by the samples of students for each year level. So, for 2017 for example, there were two sets of information on where the Years 3, 5, 7 and 9 results should be located on the NAPLAN scales established back in 2008. Both were used. 'The results of common person [horizontal] equating were checked against the results of 2017

common-item vertical equating and both sets of results were taken into consideration in finalising the scaling of the reading, spelling, grammar and punctuation and numeracy tests' (ACARA, 2018e, p. 39). The NAPLAN 2017 Technical Report (ACARA, 2018e) provides detail on the consistency of the two scaling results for each of the 16 scales, reading, spelling, grammar and punctuation, and numeracy at each of the year levels, 3, 5, 7 and 9 (pp. 40-58) and the resolution of inconsistencies between the horizontal and vertical scaling (pp. 58-60). In the process, the performance of all the link items is reviewed to identify any showing a marked difference in relative difficulty, compared with the difficulties of other items, in the link. These items are then removed from the link but retained as part of the current performance measure.

Occasionally, the equating procedures reveal what could be a deficiency in a current test. In the 2017 Year 9 reading test, the distribution of students' results, compared with previous years, was compressed at the top end, with proportions of students in Bands 9 and 10 considerably smaller than in previous years. The issue then was whether this was a real decline in the performance of high achievers or a consequence of the test poorly measuring the high achievers. Based on further investigation, ACARA's international Measurement Advisory Group agreed that the results were a consequence of poor measurement not poor student performance (ACARA, 2018e, p. 60). The solution was to adjust the distribution of results on the 2017 Year 9 reading test to match the mean and standard deviation of the results on the 2016 Year 9 reading test. This, of course, obliterated any real change in performance of Year 9 students in reading, either up or down, that might have occurred between 2016 and 2017<sup>4</sup>.

<sup>4</sup> As a matter of full disclosure, Barry McGaw, one of the members of the NAPLAN Review Panel is a member of the ACARA Measurement Advisory Group.

In 2018 and 2019 and, prospectively, in 2021 when both paper and online forms of NAPLAN are in use, the scaling process becomes more complicated because it now involves not only the common-person horizontal and common-item vertical equating, but also the use of data from different modes of assessment. The procedures used are described in the NAPLAN 2019 Technical Report for the most recent case where about 50% of students took the paper test and about 50% the online test (ACARA, 2020e, pp. 103-158).

Vertical equating for both the paper and online versions involved, as before, common-item scaling with items common in each form between adjacent tests, 3/5, 5/7 and 7/9. Horizontal equating involved common-person scaling using the secure equating test that had been administered to samples of students from 2009. This time it was given in paper form to samples of students from the populations taking both the paper and online versions two weeks later. As in earlier years, the link items were then reviewed to remove any for which there was a marked difference in relative difficulty compared with other link items. The final shifts to locate the current student performances in both the paper and online test modes in Years 3, 5, 7 and 9 on the historical NAPLAN scale from 2018, involved resolving differences between the shifts suggested by the vertical and horizontal equating.

To see if there were mode effects exerting arbitrary influences on the results unrelated to differences in student achievement, the distributions of results in 2019 for both the paper and online groups were compared with the distributions of the results for the schools involved in previous years. These comparisons were examined by ACARA's National Assessment, Data, Analysis and Reporting Reference Group, a group that has representatives from all education departments, test administration

authorities (where these are different from the department), Catholic and independent school systems/sectors and other relevant stakeholders.

This group determined that, where there were inconsistencies in the distributions, the shape of the 2019 distribution, either paper or online, should be adjusted to match that in the distribution of the results in the corresponding 2017 paper test for schools involved in the relevant group, paper or online. As with the earlier adjustment of the distribution of the 2017 Year 9 reading results to match the distribution of the 2016 Year 9 reading results, these adjustments of 2019 distributions to match the corresponding 2017 distributions similarly obliterated any real change that might have occurred from 2017 to 2019. The numbers of 2019 distributions adjusted in this way are shown in Table 14 (ACARA, 2020e, pp. 151-152).

That the distributions of 10 of 16 online scales and 8 of 16 paper scales needed adjustment indicates that achieving satisfactory horizontal and vertical equating is difficult when using two modes of assessment. The task will become easier when all assessment is in a single, online mode from 2022.

The equating will also be more secure when full advantage can be taken of NAPLAN items no longer being released and more being available for use in links, both vertical and horizontal. All links could then be based on common-item scaling without recourse to common-person scaling through a sample of students doing an additional set of NAPLAN tests.

**Table 14:** 2019 scales for which distributions were adjusted to match those for 2017

	Year 3	Year 5	Year 7	Year 9
<b>Online version</b>				
Reading	Yes	–	–	Yes
Spelling	–	Yes	–	–
Grammar and punctuation	Yes	Yes	Yes	Yes
Numeracy	Yes	Yes	–	Yes
<b>Paper version</b>				
Reading	Yes	–	–	Yes
Spelling	–	Yes	–	–
Grammar and punctuation	Yes	Yes	Yes	Yes
Numeracy	–	–	–	Yes

### Confidence in measurement

NAPLAN results are reported for individual students and in aggregate for various grouping – schools, groups of students (Indigenous, language background other than English), jurisdictions (government, Catholic and independent), state and territory and national.

There are two sources of uncertainty in NAPLAN scores – uncertainty in measurement and uncertainty in equating. Uncertainty in measurement is a consequence of NAPLAN collecting data on student achievement with relatively short tests administered on a single occasion. A parallel test, with different items covering the same curriculum domain, would be unlikely to yield exactly the same result for each student. Longer tests would yield more precise results.

Uncertainty in equating is a consequence of the common items not having exactly the same relative difficulty levels in each of the tests in which they are embedded, that is, 3/5, 5/7 and 7/9. When common-person

equating is used with all students taking the relevant NAPLAN test and a sample of those students taking the secure test from 2009, there is also uncertainty due to the sampling. If a different sample of students were used, the common-person equating would not yield exactly the same results.

How precise results are depends on how much data they are based on. Considering average results, the most precise estimates are national means. Means for larger schools would be more precise than means for smaller schools. The least precise are individual student results. The extent of uncertainty is shown in Table 15 for cases at the national mean of 495.9. The national mean is essentially a precise measure because it is based on so much data and so has virtually no uncertainty.

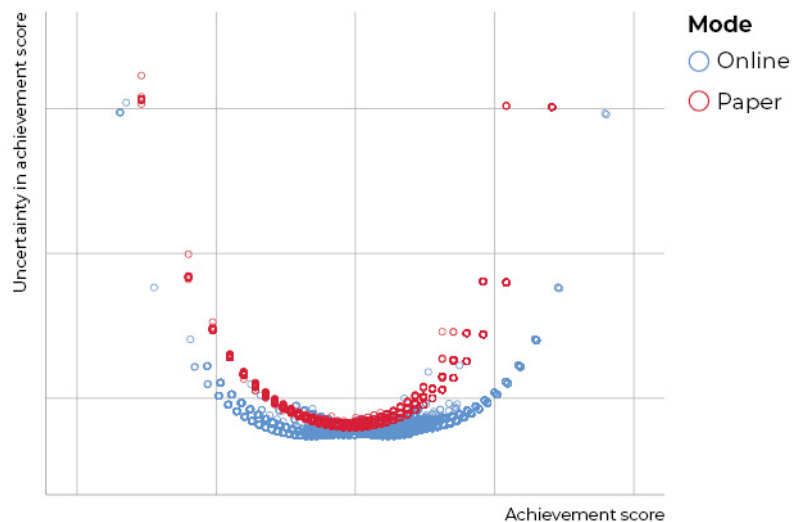
**Table 15:** Confidence ranges (95%) for the 2019 NAPLAN Year 5 numeracy scores

	Means			Individual student
	Nation	School		
		50 students	25 students	
Upper bound		513.3	519.6	539.4
Mean	495.9	495.9	495.9	495.9
Lower bound		478.5	472.2	452.4

In a school with 50 Year 5 students and a mean of 495.9, the level of uncertainty in the data means that, with a parallel test notionally administered on 100 different occasions with the same students, the means on 95 of the occasions would be expected to range from 478.5 to 513.3. On the other five occasions, the mean would be expected to be outside this range. The range from 478.5 to 513.3 is the 95% confidence interval for the mean of a school with 50 students in Year 5.

In a smaller school with 25 students in Year 5, the estimate of the mean is less precise and so the 95% confidence interval is wider. For such a school with a mean measured to be at the national mean of 495.9, it can be said with 95% confidence that the mean is between 472.2 and 519.6. For a single student measured to be at the national mean, it can be said with 95% confidence that the student's result is somewhere between 452.4 and 539.4.

The level of uncertainty in individual student results varies depending on how far the student is above or below the average. The level of uncertainty is greatest with extreme scores at the tail ends of the distribution and smallest for those in the middle of the distribution. For a student at the national mean, as illustrated in Table 15, the 95% confidence interval extends to 43.5 points above and below the measured score. For a student with an extreme score, the 95% confidence interval would extend more than 100 points above and below the student's measured score, and that is more than two NAPLAN bands.



**Figure 6:** Extent of uncertainty in student NAPLAN results with print and branching digital tests, 2018

The extent of uncertainty is greater for scores further from the mean. It would be expected to be greater for a single print test form that all students take than for a branching, digital test in which students are presented with items close to their achievement level to obtain a more precise measurement. This is illustrated in Figure 6 for both the print and branching digital forms of the 2018 Year 3 numeracy test in which students responding to the print test are shown in red and those responding to the branching digital test are shown in blue (ACARA, 2020e, p. 184).

The location of the dots in the figure are determined by the students' NAPLAN results (achievement score) on the horizontal axis and by the extent of uncertainty associated with the score on the vertical axis. The centre of the horizontal axis is located at the national mean and there the extent of uncertainty associated with individual student scores is lowest, although slightly lower for those taking the branching digital test (in blue) than those taking the print test (in red). For NAPLAN results away from the mean, the extent of uncertainty rises the further from the mean students' results lie, either above the mean (to the right) or below the mean (to the left). That increased

uncertainty occurs with both the print and branching digital tests but is less for the branching digital test. That is revealed by the blue dots being lower on the graph than the red dots for all results away from the mean. Just as Figure 5 showed how the branching, digital form measured the full range of achievements better than either a common print or digital test could, Figure 6 shows that the branching digital form measures students' achievements throughout the range with less uncertainty.

It is also important to recognise that the level of uncertainty in a simple growth measure (such as the difference between a student's successive NAPLAN results or between a school's successive mean NAPLAN results) is greater than the uncertainty associated with either of the two NAPLAN results from which the growth is calculated.

The annual NAPLAN Technical Reports provide information on the confidence bounds for those interested in examining the precision of the measurements. However reports to students and parents/carers provide only the actual scores of the student with no indication of the uncertainty of the score.



Such a level of uncertainty is not unique to NAPLAN. Patterns of uncertainty would be essentially the same for other standardised tests such as the Australian Council for Educational Research's (ACER) Progressive Achievement Tests that many schools use. The level of uncertainty would be greater for most locally developed assessments that teachers create and use because they would likely be less reliable measures than standardised ones. Reliability and validity can be increased with teacher-made assessments when multiple assessments are used, as long as 'halo effects' do not cause one result to influence another arbitrarily.

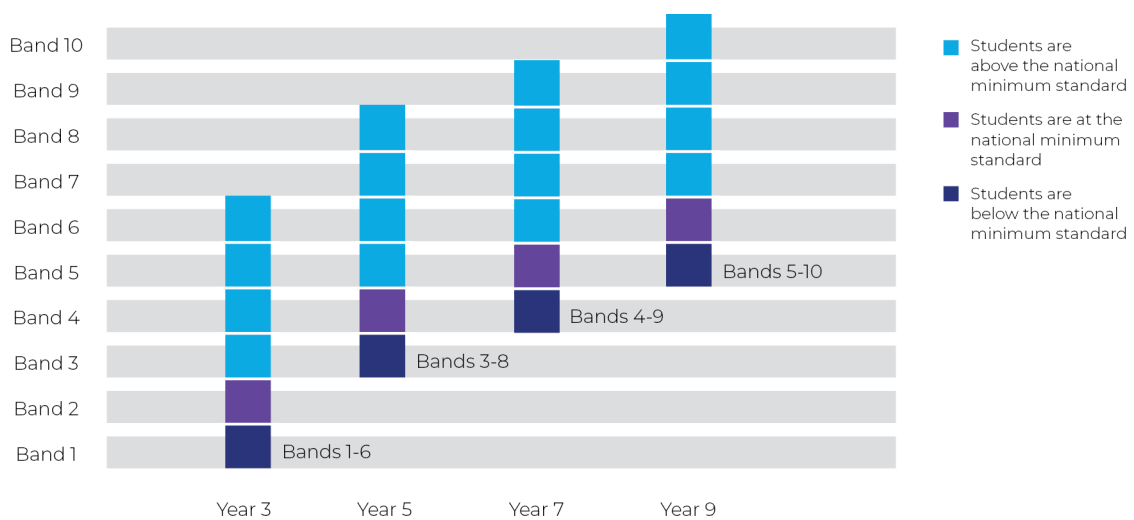
Teachers understand uncertainty in measurement since they see the variations in assessed results for individual students over time. It is why so many teachers said in submissions to the review, and in consultations that they 'triangulate' with multiple measures to obtain an increasingly stable assessment of individual students.

Uncertainty in measurement does not have a differential effect on horizontal equating since the secure tests are administered to a large sample of 800 students, large enough to obtain stable estimates of item difficulty. Uncertainty in horizontal equating is a consequence of the relative difficulties of the items in the secure test revealed in the performance of the current sample of students not being the same as those obtained in 2009. The uncertainty in horizontal equating affects the confidence with which longitudinal comparisons can be made in investigating trends over time in the results of successive cohorts of students.

The two types of uncertainty (uncertainty in measurement and uncertainty in equating) have different impacts on interpreting test results, depending on the level of aggregation of the data. For student's results, uncertainty mostly comes from measurement uncertainty. However, as the aggregation moves up, for example to the state or national averages, uncertainty in measurement has virtually zero effect and other sources of uncertainty, including equating uncertainty, become the dominating source. It will be very important to monitor the impact on uncertainty in equating when NAPLAN is fully online and the equating is exclusively common-item equating with no common-person equating, to see if the anticipated benefits accrue.

### **Establishing benchmarks**

There are single scales for NAPLAN literacy and numeracy, each with the form shown in Figure 7 (ACARA, 2019b, pp. vi). The vertical equating, discussed above, enables the students' results in Years 3, 5, 7 and 9 all to be located on the same scale. The scale has ten bands but, as can be seen in Figure 7, only part of the range is used at each year level.



**Figure 7:** NAPLAN assessment scale

As described in Chapter 2, there is a National Minimum Standard (NMS) set on each scale for each year level. The NMS is defined in the following terms.

The second lowest band on the achievement scale reported for each year level represents the national minimum standard expected of students at that year level. The national minimum standard is the agreed minimum acceptable standard of knowledge and skills without which a student will have difficulty making sufficient progress at school (ACARA, 2019c, pp. vi).

There is no indication of how these benchmarks became ‘the agreed minimum acceptable standard’ in each domain but it is clearly a matter of professional judgement and consensus.

As also described in Chapter 2, there are benchmarks set on the scales for the international surveys of student achievement – Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) with which comparisons were made with

NAPLAN. The discussion in Chapter 2 focused on similarities and differences in trends over time in the percentages of students performing at or above the benchmarks on the various scales. The international scales have more than one benchmark, as indicated in Chapter 2.

The actual percentages below the lowest benchmark in the latest assessments in reading/literacy and numeracy/mathematics in each of the surveys are shown in Table 16, where it is clear that fewer students fail to reach the NAPLAN benchmarks than fail to reach the other benchmarks. This could be because the students are better prepared for the NAPLAN tests because of their connection with the Australian Curriculum, or engage with them more because they are domestic census tests rather than international sample surveys. The other obvious possibility is that the NAPLAN NMS benchmarks are less demanding than the international ones. That is even more likely given that all students who did not sit the NAPLAN tests because they were exempt are counted as below NMS. The percentage of those who sat and are below NMS is therefore smaller than the percentages shown in Table 16.

**Table 16:** Percentages of Australian students below minimum standards benchmarks

	NAPLAN			PIRLS	TIMSS		PISA
	Year 3	Year 5	Year 9	Year 4	Year 4	Year 9	15-years
Reading/literacy	4.1	5.3	8.2	13.0			19.0
Numeracy/mathematics	4.5	4.6	4.0		9.0	11.0	19.6

The 1996 National School English Literacy Survey provided earlier data on the performance of Years 3 and 5 Australian students in reading and writing (Masters and Forster (1997b). The then federal Minister for Education commissioned Masters and Forster to set minimum performance standards as benchmarks for Years 3 and 5 students. They did this with samples of students' work from the national survey, obtaining judgement from teachers, as well as from literacy and numeracy specialists, on whether samples of students' work were above or below the standard they expected of students at those year levels. This enabled Masters and Forster to locate the benchmarks on the reading and writing scales developed in the national survey. The conclusion was that 27% of Year 3 students and 29% of Year 5 students did not meet the relevant benchmarks (Masters and Forster, 1997a, p. 15).

Masters and Forster's work and the international surveys all beg the question of whether the NAPLAN National Minimum Standards are set too low. There is work underway to set additional higher benchmarks on the NAPLAN scales from 2021 for 'proficient' and 'highly proficient' performance (ACARA, 2019b, pp. 10-11).

## Inclusiveness of the tests

The NAPLAN testing program is designed as a census assessment in which all students are intended to participate, though parents/carers can withdraw their child. There are three reasons why a student may not participate in some or all of the NAPLAN tests. They are:

**Exemption** – Students with a language background other than English, who arrived from overseas less than a year before the tests, and students with significant disabilities.

**Withdrawal** – Students withdrawn by their parent/carer based on religious beliefs or philosophical objections to testing.

**Absence** – Students not present at school because of an accident or mishap or by choice (ACARA, 2018e, pp. vii-viii).

The reasons for non-participation at a national level are shown in Table 17 for NAPLAN 2017, the last year in which only the print form was used (ACARA, 2018e, pp. 59, 123, 187, 251). The rate of exemptions and withdrawals was generally consistent over all the year levels tested but the rates of absence on the days of testing were higher in the secondary years and particularly at Year 9.

**Table 17:** Percentages of non-participating students in NAPLAN 2017 tests

	Reading	Writing	Language conventions	Numeracy
<b>Year 3</b>				
Exemptions	1.9	1.9	1.9	1.9
Withdrawals	2.8	2.9	2.8	2.7
Absences	2.3	2.4	2.2	2.7
<b>Year 5</b>				
Exemptions	1.9	1.9	1.9	1.8
Withdrawals	2.3	2.3	2.3	2.2
Absences	2.3	2.4	2.2	2.8
<b>Year 7</b>				
Exemptions	1.8	1.8	1.8	1.7
Withdrawals	2.1	2.1	2.0	2.1
Absences	3.5	3.4	3.2	4.0
<b>Year 9</b>				
Exemptions	2.0	2.0	2.0	2.0
Withdrawals	2.7	2.6	2.6	2.7
Absences	6.0	5.8	5.6	6.6

**Table 18:** Participation rate (%) in NAPLAN in 2017

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Year 3	97	95	93	95	93	95	94	88	95
Year 5	97	95	93	96	94	95	94	88	95
Year 7	97	95	91	96	94	94	95	86	95
Year 9	95	91	87	94	90	90	90	79	91

Participation rates vary across year levels and states and territories, as can be seen in Table 18 for NAPLAN 2017. Participation rates are lower for Year 9 than Years 3, 5 and 7 and they are lower in the Northern Territory, South Australia and Queensland than in the other states and territories. Non-participation is not randomly distributed across student groups, with low performing students (students with lower prior achievement scores) more likely to be absent from the tests than their counterparts (Centre for Education Statistics and Evaluation, 2016). As the test results are not missing at random, this can bias the state and territory mean scores. It would be good for jurisdictions to investigate students' reasons for absence and seek to reduce the current levels.

### Students with disability

Adjustments to NAPLAN testing are available for students with a disability. These include assistive technology that does not 'compromise a student's ability to independently demonstrate the literacy or numeracy skills that are being assessed through the NAPLAN tests' (for example, text-to-speech), alternative questions (for example, audio presentation, adjustments in font size, variations in colour), support persons (for example, a scribe for the writing test or a reader for other tests) and extra time and rest breaks (where these are part of the student's regular teaching and learning experience) (ACARA, 2020f). There is also provision for some students to be withdrawn.

In cases where the severity or complexity of a student's disability does not allow the student to participate in NAPLAN, or where a student is from a non-English-speaking background and arrived in Australia less than one year before the tests, students can be exempted from one or more NAPLAN tests (ACARA, 2020i).

One submission noted a marked improvement over time in the inclusiveness of the testing arrangements.

As a teacher I was previously against the Basic Skills Test in NSW as I taught in a poor and culturally diverse community. I have been impressed with the way assessments have continually been made more inclusive. (Education expert)

Perhaps because they are involved with adjustments for students taking end-of-secondary school examinations and other assessments, secondary schools were said to be better than primary schools in accommodating students' needs, – 'Secondary schools are generally much more aware of the accommodations available for students with disabilities who sit NAPLAN.' (Disability group representative)

Parents/carers and representatives of agencies that provide support for children with disabilities, particularly learning difficulties, stressed the importance of students' participation to obtain an external assessment of the students' progress.

Parents/carers of children with disabilities don't have many sources to compare their child's performance with other similar students. We don't want to lose the value of NAPLAN for that group of people. (Parent/carers association)

A lot of parents/carers come to external organisations for an assessment after Year 3 NAPLAN, because, if their child doesn't do well at the national level, it "sets off alarm bells". There are also more referrals after Year 7 NAPLAN, because primary schools don't generally pass on information about students with disabilities to secondary schools. That usually falls to the parents/carers. (Disability group representative)

Some parents/carers complained that the decision on withdrawal was effectively taken from them by school principals.

Some students with disabilities have been asked by schools to not participate in NAPLAN. This is due to the perception that these students will have a negative effect on a school's NAPLAN results. (Disability group representative)

Many parents/carers are told it would be best for their child not to do NAPLAN. It is the parent's right to have their child do NAPLAN. Schools don't dissuade in writing, but they do tell children not to participate. (Parent/carers' association)

Kids don't discriminate but, when schools identify children with disability and keep them out of NAPLAN, it is noticed by the other students and it begins the process of discrimination. (Parent/carers' association)

On the other hand, it was reported that, if school funding is based on evidence of student need, participation can be encouraged – 'If funding is attached to NAPLAN data, some schools encourage students with disabilities to attend tests so that they can attract higher funding.' (Disability group representative)

## **Aboriginal and Torres Strait Islander students**

Most Aboriginal and Torres Strait Islander students live in capital and regional cities. A minority live in remote communities, but there is evidence that students from both settings may face some problems with NAPLAN. Exclusion from the tests is one.

Comparing schools through NAPLAN is concerning. This has resulted in a negative effect of some Aboriginal and Torres Strait Islander students being asked to stay home on the day of the tests to improve school results. (Aboriginal and Torres Strait Islander representative body)

There are concerns about the validity of the tests for many Aboriginal and Torres Strait Islander students. For some, it is the lack of recognition of students' ability in languages other than Standard Australian English (SAE) – 'English as a second language is discussed in the negative, rather than the benefits of being multilingual.' (Aboriginal and Torres Strait Islander representative body)

Children who grow up with English as first language still have difficulty. The assumption is that, because they don't speak their own language, they didn't qualify to be recognised as students who would need extra assistance. (Aboriginal and Torres Strait Islander representative body)

Students in remote communities do not speak SAE at home, in the playground or in the classroom. Regional and metro children who speak Indigenous English as another Language or Dialect (IEAL/D) at home still need to learn English for peer interaction at school. This is not the case for remote Aboriginal and Torres Strait Islander students. We already know that the SAE nature of NAPLAN excludes these students. Making these stats available to the public, often used by media, perpetuates the myth of inferior Aboriginal and Torres Strait Islander student ability. (Online submission respondent)

For others, the lack of validity of the NAPLAN tests lies in the exclusion of Aboriginal and Torres Strait Islander knowledge, – 'There is research by other Aboriginal and Torres Strait Islander education researchers that shows that NAPLAN falls short of reporting some really important attributes in how Aboriginal students see the world.' (Aboriginal and Torres Strait Islander representative body), and

'The Stronger Smarter Institute released results that showed Aboriginal and Torres Strait Islander students outperformed non-Aboriginal and Torres Strait Islander students

in environmental knowledge. This reflects Aboriginal and Torres Strait Islander student capabilities.’ (Aboriginal and Torres Strait Islander representative body)

Nevertheless, there was a strong view that assessment in basic literacy and numeracy is important for Aboriginal and Torres Strait Islander students, – ‘Literacy and numeracy are incredibly important from the outset.’ (Aboriginal and Torres Strait Islander representative body), ‘NAPLAN has a “very important message of giving a good start to get a good finish”. It is important to make sure all our kids get a good start.’ (Aboriginal and Torres Strait Islander representative body), and ‘Our whole community suffers when we are not creating people who will contribute to their communities in a constructive way. We want NAPLAN to benefit our kids so they can contribute to society.’ (Aboriginal and Torres Strait Islander representative body)

### **Cultural and language diversity**

Migrant students also have the advantage of speaking more than one language but, in taking the NAPLAN tests while still acquiring Standard Australian English, can have their academic achievements underestimated.

There is a general consensus amongst Teachers of English to Speakers of Other Languages (TESOL) scholars that standards-based assessment regimes are not inclusive of students learning English as an Additional Language or Dialect (EAL/D) because they operate from a monolingual paradigm that fails to acknowledge how the language and literacy practices of multilingual learners differ. (Written submission response: educational expert)

Our national curriculum and assessment frameworks must truly recognise the linguistic diversity of Australia’s student population by providing distinct learning and assessment pathways that are tailored to our students’ English language learning needs. (Written submission response: educational expert)

### **Similarity to other tests used in schools**

NAPLAN tests capture students’ performance on a single occasion in Years 3, 5, 7 and 9 so they give a limited snapshot of students’ development. Teachers’ regular observations and assessments of students work add a richness to the view, but they lack the comparative perspective that comparable assessments across the system, state and nation can provide.

Many schools use other external, standardised assessments to obtain further, objective information on their students’ achievements. The most commonly reported as used in schools are Australian Council for Educational Research (ACER’s) Progressive Achievement Tests (PATs) which are available to assess Early Years (mathematics and reading in the first two years of schooling), reading, vocabulary skills, spelling, punctuation and grammar, mathematics, science, and STEM contexts (inquiry and problem solving in the domains of science, technology, engineering and mathematics). All the tests are available online and in print except for PAT-R Spelling which is available only in print and PAT Early Years and PAT STEM contexts which are available only online. The PATs typically cover the range from early primary to Year 10, are mapped to the Australian Curriculum and provide an external reference for interpreting levels of performance through norms that reflected the distribution of performances in a relevant, reference population (ACER, 2020a).

Other assessments nominated as in use in schools include:

- The PM Benchmark Reading Assessment Resources (Nelson, 2020).
- ICAS Assessments (UNSW Global, 2020b).
- Reach Assessments (UNSW Global, 2020a).
- Academic Assessment Services (2020).
- York Assessment of Reading for Comprehension (GL Assessment, 2020).
- PROBE2 Reading Comprehension Assessment (Parkin & Parkin, 2011).

These tests produce results with the same kind of measurement uncertainty that NAPLAN does. They avoid uncertainty due to equating over time (horizontal equating) because they do not attempt to monitor system-level changes over time. One reason that some schools prefer these other standardised tests to NAPLAN is that the results are available much more quickly, – ‘We place greater trust in the ACER tests where we can get data straight away.’ (Member of the NAPLAN Review Practitioners’ Reference Group)

Another reason for preferring other tests to NAPLAN is that schools have control over the use of results. NAPLAN results are made public; the others need not be, – ‘The advantage of PAT testing is that marking “can be kept in-house”.’ (Member of the NAPLAN Review Practitioners’ Reference Group)

Some schools prefer not to use standardised tests at all, – ‘We don’t use a lot of standardised tests or do many tests. We use anecdotal notes, observation and moderation against work samples to assess students.’ (Member of the NAPLAN Review Practitioners’ Reference Group)

Our main data source is from formative assessment and conferencing with students. We also use focus groups. We look at where students are at against achievement standards and what they need. When we get NAPLAN results, we often see big discrepancies between assessments. (Member of the NAPLAN Review Practitioners’ Reference Group)

Submissions and consultations regularly referred to schools ‘triangulating’ data from different sources to best understand students’ achievement levels. There is, however, no straightforward way to combine the information and, apparently, little attempt to integrate the NAPLAN results in schools’ reports to parents/carers and students.

Comments received as part of this review indicated that NAPLAN results are provided to parents/carers and students in different ways. Many send the individual reports home in sealed envelopes. Some add a covering letter with general comment on NAPLAN and, usually, an invitation to discuss the report with the child’s teacher though one secondary school department head added, ‘To be honest, I think many teachers themselves wouldn’t be completely sure how to interpret the data “as is”’. Whatever the communication strategy, little link seems to be made between NAPLAN results and schools assessments. That may not be surprising with NAPLAN results being returned to schools, months after the testing. (The ACT gets around this problem by distributing interim results for all but writing to schools as soon as available and before the end of Term 2.)

Some comments referred to triangulation as desirable but not straightforward, – ‘We use ACER tests for reading, maths and spelling. We’d like to try to triangulate with NAPLAN data but the results don’t necessarily work together’ (Member of the NAPLAN Review Practitioners’ Reference Group).



There is currently a lot of “piecemeal” data analysis occurring. Matching NAPLAN assessment data to other assessment data would be a powerful student data mechanism. Ideally, this data could later be matched with teacher judgement. (Principals’ associations)

The term ‘triangulation’ seemed to be used quite loosely in much of the comment in the consultations and submissions to describe only a process in which several pieces of information are borne in mind as an overall judgement of a student’s performance is formed. That raises all the regular questions about validity and reliability of assessment – ‘How replicable would one person’s use of the data or other evidence be?’ ‘Would a different teacher reach the same overall judgement?’ This is not to suggest that ‘triangulation’ be abandoned as a term or as a practice, it is to suggest that it needs to be made as rigorous as possible and as collaborative as possible.

There is a new Australian initiative that is intended to make this task much easier for schools. It is the Online Formative Assessment Initiative being undertaken by the Australian Curriculum, Assessment and Reporting Authority (ACARA), Education Services Australia (ESA) and the Australian Institute for Teaching and School Leadership (AITSL) which:

aims to provide Australian teachers with innovative assessment solutions that integrate resources, data collection and analytical tools in one ‘ecosystem’ that is easily accessible, interactive and scalable to meet future needs. The initiative will give teachers the tools, flexibility and professional learning they need to plan teaching that will work best for the students in their classroom. It will also give students more insight into their learning and better understanding about next steps to improve progress.

...

The ecosystem will help teachers who want to use online formative assessment identify where their students are in their learning and then work with students on their next learning steps by identifying and using effective teaching practices and quality resources. The system will offer access to quality assessments and digital resources that are aligned to the National Literacy and Numeracy Learning Progressions and the Australian Curriculum.

The ecosystem will also help teachers bring together information about student learning from a range of tools or resources that they might already be using, to create a coherent view of progress that can be shared with the students and parents (ACARA, ESA, AITSL, 2020).

Students’ NAPLAN results could be one piece of evidence placed into this ‘ecosystem’ and there it could be more readily connected with the other information that the teacher holds.

## Summary

The NAPLAN tests in literacy, language conventions and numeracy, developed in 2007 and used for the first time in 2008, were the product of a national collaboration created by the ministerial council to build upon separate state and territory assessments that had been developed and implemented over the previous two decades. The new national tests were based on new national Statements of Learning for English and the national Statements of Learning for Mathematics until 2016, after which they have been based on the Australian Curriculum.

The first major change came with the introduction of a computer delivered, digital form of the tests in 2018 that was used by just over 15% of schools while the rest used the established print form. In 2018, over 50% of schools used the digital

form. It was expected that all schools would have switched to the digital form by 2021. With the 2020 NAPLAN testing abandoned because of the COVID-19 virus, full adoption of the digital form is now anticipated in 2022.

Using the digital and print forms of the NAPLAN tests in parallel meant that the digital form was required to match the print form with the consequence that the digital form could not exploit the full capacity of electronic delivery to use more innovative test items. That constraint will be removed when all schools use the digital form.

There is, however, one constraint in the print form that has already been removed for those schools using the digital form. With the print form, all students answer the same set of questions and spend some of their time answering questions that are either too difficult or too easy for them and so provide little information on how well the students are performing. Computer delivered tests can be similarly inflexible but, with NAPLAN, the digital forms are adaptive. One third of the way through the literacy and numeracy tests, with students' responses having been marked as they answered, the students are branched to more or less complex items for the next third of the test. Then, at two thirds of the way through the test the students are branched again to more or less complex items. (See Figure 2, p. 60.)

Better matching the test items to the students' achievement levels in this fashion provides better coverage of the range of student's achievements at the high and low ends of the distribution than can be achieved with a single print form that all students take. (See Figure 5, p. 67.)

The branching digital form also provides measures of student achievement with less uncertainty than the common print form, particularly of high and low achievers. (See Figure 6, p. 72.)

National Minimum Standards are set on the NAPLAN scales for each year level. The percentages of students performing below those levels are smaller than the percentages of Australian students performing below the minimum standards set in all of the international surveys of student achievement in which Australia participates. Australian students may do better on NAPLAN because it better fits with the Australian Curriculum but it could equally be that the National Minimum Standards are set too low on NAPLAN.

The NAPLAN tests are intended to be taken by all but a small number of students who are exempt because they have significant disabilities or have recently arrived in Australia and have a language background other than English or who are withdrawn on the request of their parents/carers on religious or philosophical grounds. There are more students who do not participate by not attending on the NAPLAN testing days. Participation rates are better at the primary level than at the secondary level and they also vary across states. The absences are not random. It is poorer performers who are most likely to fail to participate, and that risks bias in jurisdiction and system/sector means.

NAPLAN provides one piece of information about students, alongside many other pieces of information that schools and teachers have. None of these other pieces is likely to be more reliable or valid than NAPLAN; but together they can all contribute to a richer, and more valid and reliable, picture of the student's achievement and progress. The ways in which the pieces of information are combined is what teachers and principals call 'triangulation', however, this could be made more rigorous and systematic.

# Chapter 5: Quality of the NAPLAN writing test

The NAPLAN writing test has been part of Australia’s national census testing program, dating back to 2008. This chapter presents an overview and critique of the test. Areas of focus include: the writing prompts; the scoring rubric; the conditions under which students undertake the test, offered in 2019 in two modes (computer-based and pen and paper based); and the alignment of the test to the Australian Curriculum: English. At the time of writing, the Australian Curriculum is under review, and work is underway on the general capabilities and the development of National Literacy and Numeracy Learning Progressions. The Terms of Reference for the current Australian Curriculum Review indicate the intent to:

revisit and improve where necessary, the learning continua for the general capabilities with reference to current research, in particular: replace the learning continua for literacy and numeracy with Version 3 of the National Literacy and Numeracy Learning Progressions and use the progressions to inform refinements to the Australian Curriculum in English and Mathematics, as well as review the literacy and numeracy demands of content in the other learning areas (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2020l, p. 4).

An international overview of the assessment of writing in selected countries was also undertaken to inform the review and is provided in Appendix 4 as supplementary information. A distilled set of NAPLAN writing test issues to be resolved and recommendations on dealing with them are presented in Chapter 7. When the below National Minimum Standards figures are stated in this chapter, these exclude exempt students.

## Key points:

- The testing of students’ writing proficiency has been a long-contested site in Australia and internationally, suggesting the complexity of the domain being assessed.
- Stakeholder consultations presented a sustained thread of dissatisfaction about the content of the writing tests. The main issues are: the prompts, the choice of forms, the criteria and the related range of scores, and the conditions under which students write.
- The writing test has had unintended effects on how writing is taught; and students’ writing for NAPLAN was frequently described as formulaic.
- Writing data from the NAPLAN National Reports (Table 19, Table 20 and Table 21) show a picture of young people reaching Year 9 without achieving writing proficiency. Concentrations of performance at below National Minimum Standard (NMS) are higher for students in regional and remote areas. The difference in performance between males and females is significant and has been evident each year since 2008.

## Key points continued

- The data indicate that writing has not improved since 2011.
- Teachers' professional judgement was repeatedly referred to as a critical missing element in the NAPLAN writing test beyond teachers' involvement in scoring as part of marker panels.
- The mode of the writing test for students in Year 3 should be pen and paper based; for students beyond Year 3, the mode should be computer-based.
- There is a widely reported lack of confidence in Automated Writing Evaluation (AWE), especially for authorial aspects of writing.
- Those students in Years 5, 7 and 9 who have prior opportunities to develop typing fluency and word processing skills by regularly using a keyboard are better placed to produce sustained writing online.
- The explicit teaching of keyboarding skills and monitoring typing proficiency beyond Year 3 are essential in developing students' writing skills and for equitable participation in computer-based writing assessments.
- How a country tests writing reflects interrelated decisions about: the purposes of testing; the stages of schooling to be included; the curriculum domains and related forms of writing to be tested; how the writing is to be scored (including criteria, judgement method, human and machine scoring); the role of the profession; quality assurance processes including online moderation; intended uses of the reported results, and to whom and how they are released. All these matters are central to a decision about whether a test is fit-for-purpose.

## NAPLAN writing test

International testing programs (for example, Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA)) have gathered information about reading literacy, mathematics and science. However, the domain of writing has not been included to date. A recent report from the United Nations Educational, Scientific and Cultural Organisation (UNESCO) (2019) acknowledged that inter-country assessment of writing is in its infancy. The report identified that assessing writing skills in domain areas is not well-advanced, generally lying beyond the scope of large-scale learning assessments. Writing is characterised as a:

...foundational skill required for communication, future learning and full participation in economic, political and social life as well as in many aspects of daily life. In a digital age and in the context of a knowledge economy, personal and social communication is increasingly conducted in written text, including through mobile phones and social media. Assessing writing skills or the use of them to measure domains, such as creativity, curiosity and the appreciation of culture, also generally lies beyond the scope of Large-Scale Learning Assessments (UNESCO, 2019, p. 42).

Contestation surrounding the direct assessment of writing is not new. Humphry and Heldsinger (2019) suggest that the current lack of agreement about best practice in assessing writing has occurred

'[p]erhaps because of the complex and multi-faceted nature of writing' (p. 3). Added to this are competing and strongly held views about assessment validity in the case of writing, and specifically the nature and scope of evidence requirements; rating approaches (holistic scoring, analytic scoring) and how markers should apply criteria, standards and numeric scales, separately and in combination, as components of valid assessment of writing performance (Messick, 1994; Wyatt-Smith & Adie, 2020). Added to these are thorny issues about the conditions under which students are expected to write, specifically the time needed for authentic writing processes, the roles of human scorers and automated writing evaluation, and the conditions under which teacher judgement can be made dependable (Harlen, 2005a, 2005b). These issues are addressed later in this chapter.

Some researchers have reported that the typical matrix design of criteria and analytic rubrics poses a threat to validity (Sadler, 2009). They suggest that criteria statements alone do not guarantee high inter-rater reliability or overall accuracy of scoring (Delandshere & Petrosky, 1998; Wilson, 2006 cited in Rezaei & Lovorn 2010) and point to how criteria can trigger 'pronounced rating tendencies of a form that would usually be interpreted to indicate a halo effect' (Humphry & Heldsinger, 2014, p. 253), discussed later in this chapter.

Three main types of writing are identified for use in the NAPLAN writing test. These are imaginative, persuasive and informative texts. To date, the latter category has not been used in the writing test, though there is provision for it to be included in future tests (ACARA, 2017). Informative writing is arguably the most common and important genre, both in professional and business writing. Perelman (2018) drew on the oft-reported link between what is tested and what is taught, asserting that 'not testing informative writing devalues it in the overall curriculum' (p. 7).

Each year, the writing prompt is the same for all children in Years 3 and 5, with a different prompt offered to young people in Years 7 and 9. The form or genre is common across the testing year levels. Since 2008, the narrative and persuasive forms have been set (narrative: 2008, 2009, 2010, 2016, 2019; since 2011, the prompts have been for persuasive writing every year except 2016 and 2019) (ACARA 2017). Readers are also advised to see the section on Writing in Chapter 2.

The recognised features of the imaginative, persuasive and informative texts are described below. Each text type has a recognised primary purpose, use in social contexts, recognisable structural features and associated linguistic characteristics.

**Imaginative texts** — texts for which the primary purpose is to entertain through their imaginative use of literary elements. They are recognised for their form, style and artistic or aesthetic value. These texts include novels, traditional tales, poetry, stories (also known as narratives), plays, fiction for young adults and children including picture books and multimodal texts. ... For a NAPLAN writing test students may be asked to write a story that is centred on an idea, tension or conflict, and use a structure that has an orientation, a complication, and a resolution.

**Persuasive writing** — texts for which the primary purpose is to put forward a point of view and persuade a reader, viewer or listener. They form a significant part of modern communication in both print and digital environments. They include advertising, debates, arguments, discussions, polemics and influential essays and articles. ... A NAPLAN writing prompt of this text type is constructed to allow students to convince the reader to adopt a given point of view or urge the reader toward a specific action.

**Informative writing** — texts for which the primary purpose is to inform. Informative writing includes explanations and descriptions with the express purpose of informing the reader. It is one of the most commonly used writing forms and is central to learning across the curriculum. ... A NAPLAN writing prompt of this type either provides the students with the necessary information or requires students to have sufficient content knowledge of the topic for them to be able to demonstrate their writing skills (ACARA, 2017, pp. 15-16)

Regarding purpose, the NAPLAN Assessment Framework (ACARA, 2017) indicates that the NAPLAN writing task is designed to assess the accurate, fluent and purposeful writing of either an imaginative or persuasive text in Standard Australian English. Officially, the 'NAPLAN writing test complements the NAPLAN conventions of language test assessing spelling, grammar and punctuation within the context of writing' (ACARA, 2017, p. 14). The assessment framework also indicates that the test aligns with the Australian Curriculum: English 'through a focus on three central types of texts that are essential for students to master if they are to be successful learners, confident and creative individuals, and active and informed citizens: persuasive, imaginative and informative' (ACARA, 2017, p. 15).

## Critique of the NAPLAN writing test

Writing is arguably the most complex performance in the three domains currently assessed in NAPLAN. It is also the domain that has attracted consistently negative comment throughout the stakeholder consultations. In the words of one respondent, "writing can't be easily fixed" (Education expert). There was a sustained thread of dissatisfaction about the writing test, with the concerns being common

across Years 3, 5, 7 and 9. NAPLAN writing results have also attracted considerable negative comment from national and state media outlets.

While there is widespread recognition among those participating in the review about the value of assessing writing in schooling, the NAPLAN writing test received criticism across stakeholder groups. While there were those who proposed that the writing test be terminated, there were also views that it should be retained, though in a redesigned form. Those taking this stance referred to how the test generated information valuable to schools and that this was not otherwise available, especially for monitoring and comparative purposes.

Significant concerns were raised about how the test is designed, implemented and reported. While the review panel heard some variation in the intensity of the views, the common thread was that overall, the NAPLAN writing test does not support students to produce excellent writing; in its current form, is not highly valued by teachers and school leaders; is not well-designed, impacts negatively on how writing is taught in the classroom; and leads to narrowing of students' literacy learning. It was frequently mentioned that students produce formulaic writing for NAPLAN. It was also common throughout the consultations to hear that the writing test is having a negative impact on children's and young people's enjoyment of writing, their creativity, and opportunities to express imagination in writing. An additional claim, often repeated, was that the test has the effect of suppressing the quality of the writing students could demonstrate at the high-end of performance in favour of attempts to deliver writing to fit 'the formula' – "NAPLAN has an effect on the "joy" of writing" (Parents'/carers' association); and "The richness of writing has been lost. 'Cookie cutter' writing is being produced" (Subject association).

## Formulaic teaching of writing and teaching writing as formulaic

The repeated observation was that NAPLAN has had unintended effects on writing pedagogies. Respondents in the consultations, including those who had been involved in NAPLAN scoring, characterised writing produced for NAPLAN as formulaic. The potential for rich writing pedagogy was talked about as being reduced to students composing paragraphs to a formula. More than this, the lesson for students was that there is a set formula for producing quality writing. In the words of one respondent, “Some students learn a piece of writing and reproduce it” (School system/sector), with another commenting, “The writing test tends to be quite formulaic and the responses seem to be quite formulaic as well” (Member of the NAPLAN Review Practitioners’ Reference Group). This observation is consistent with reported downward pressure in some sites on teachers’ use of mock tests or rehearsals for NAPLAN writing that consumed considerable teaching time. The space for creativity or imagination, and opportunities for what some teachers have referred to as being playful with language appears to be shut down.

Stakeholders also highlighted factors external to the test itself that impacted Years 3, 5, 7 and 9 and non-NAPLAN Years (2, 4, 6 and 8). They reported growth in tutoring businesses with some parents/carers paying tutors to prepare students for the writing test and other domains of the test. Also identified was the active role of private testing companies that offered schools opportunities to sit NAPLAN-like tests or simulations, with student work scored externally and reported back to the school. This scheduling of local versions of NAPLAN, like testing and reporting, had been added to some schools’ own programs of assessment and, in some cases, they had become ‘normalised’ over time.

There was a range of responses addressing test preparation and time spent on building student readiness for writing as a solo performance. At the one end of the continuum is the position that preparing for NAPLAN writing interrupted the school’s curriculum program delivery. This occurred as teachers and students stopped planned learning to focus instead on imaginative or persuasive writing in the weeks leading up to the testing period, with practice sessions where students sat mock writing tests. These provided students with the experience of composing under restricted time conditions, with no scaffolding and no access to material and human resources.

Where this displacement of curriculum occurred, NAPLAN test preparation became the proxy curriculum and teaching writing was reduced to a dominant focus on the structure of the writing, the formula for producing the narrative or the persuasive form, with some commentary about maximising the score on writing through ‘gaming’ the criteria. In commenting on classroom preparation of students for NAPLAN writing, one respondent said “You’re forensically taking out different things, and you can formulaically produce a response that maximises the result” (Member of the NAPLAN Review Practitioners’ Reference Group), with another commenting that “the writing assessment is formulaic and focuses on structure over content” (Member of the NAPLAN Review Practitioners’ Reference Group). Especially for students in Years 3 and 5, the requirement to produce writing independently was described as ‘alien’, that is, missing the types of scaffolding including time for planning, drafting and editing, and feedback that teachers reported to be routine in the classroom.

Stakeholders’ comments showed a common interest in how the writing test in its entirety (choice of forms, prompts, criteria and scores) could maintain a relevance

to classroom practice and the learning of individual students and sub-groups. There were numerous calls to provide greater scaffolding or supportive framing to be built into the test design evident in the illustrative segments below – ‘Primary students are never expected to write under these conditions unless they are preparing for the test. Learning should never be about preparing for a test. (Respondent to the online survey).

A student should be permitted to submit three drafts of a piece of writing. They should be given three times to complete each stage of the process with access to reference materials like dictionaries to improve their work. (Respondent to the online survey)

NAPLAN could incorporate a longer-term written component, possibly across 1 term or several weeks, in which quality of argument based on thorough background research is prioritised over the ability to write a lot in a highly stressful environment. (Respondent to the online survey)

Respondents also commented on how the NAPLAN writing test results and the reports specifically did not play an important role in teacher discussions with parents/carers about student progress or achievement. Some teachers indicated that they had no recollection of a parent initiating a discussion with them about their child’s writing results, though, as noted elsewhere, primary teachers reported being aware of high schools asking parents/carers to provide NAPLAN results for enrolment and related screening purposes for high school enrolment.

While it is required practice for schools to send home the NAPLAN reports, very few references were made to teachers and school leaders arranging to meet with parents/carers about the data in the reports.

Talk with parents/carers about NAPLAN writing results being used for goal setting and informing next step teaching appears to be limited, reflecting the teachers’ widely reported perception that the writing test arrives back in the school too late to be used for diagnostic purposes. This observation is consistent with earlier research reporting that NAPLAN results were not regarded by teachers as having a clear ‘feedforward’ function, unless system/sector and school leadership enabled this to happen using a range of strategies including target setting and a related focus on teachers’ assessment literacy. While some jurisdictions have investigated NAPLAN writing results and school reporting of student achievement in English, discussion about how the two relate at the level of cohorts and individual young people did not feature as part of routine parent/carer and teacher communication about students’ progress in writing. The limited profile given by teachers and parents/carers could reflect a perceived significant difference between the NAPLAN writing assessment and how writing is assessed in classrooms, as mentioned above. This is also evident in the responses below, especially for students from a range of cultural and linguistic backgrounds – ‘The marking schema is seen by many to be formulaic and the criteria may not reflect how writing is generally assessed in classrooms’ (Respondent to the online survey); and ‘It has been suggested that the marking criteria should include a greater recognition of the genre characteristics as part of quality writing.’ (School system/sector).

### **Factors internal and external to the test**

Overall, teachers, school leaders and union representatives talked about how, in its current form, the NAPLAN writing test was not regarded as a world class test. Concerns included that it lacked



authenticity or relevance for students and did not reflect adequately the key aspects of writing pedagogy and the range of valued characteristics of writing that should be covered in a high-quality writing test. It was characterised as having significant design limitations. These included the choice of prompts, the limited range of forms or text types, the marking guide, including the stated criteria and the accompanying numeric scores, the limited attention to writing purpose and audience, and the 'alien' conditions under which the students were expected to produce their piece of writing.

According to teachers and school leaders, in the main students did not see NAPLAN writing as relevant to how they learn about 'good writing'. There was consensus that using students' 'on-demand writing' in restricted time conditions means that students did not have the opportunity to demonstrate their best writing, irrespective of the mandated prompt for the year.

Further, as a limited snapshot of writing on demand, the widely reported comment was that the NAPLAN writing reports had little, if any, diagnostic utility. As mentioned earlier, reports arrived back in the school too late to inform targeted interventions at the level of the whole class, sub-groups in a cohort and individual students. A frequently reported view was that the writing assessment did not tell teachers anything that they did not already know. For those espousing this view, the value of NAPLAN writing results was that they served to confirm teachers' own assessments of student writing.

This is a point regularly made, suggesting that there are few, if any, surprises for the teacher in the writing reports showing ordering of students in the class. What NAPLAN can add is comparative information about the level of performance in other schools with comparable students. There were also repeated references by teachers and school leaders to the practice of triangulating data where different types of

assessment evidence are considered, to see patterns and possible areas for intervention. Very few of these related to NAPLAN writing data. Where they occurred, there were some references to spelling for school target setting – 'We pin things down to specific areas. For example, if spelling dips down in comparison to previous years [...]. We've done interventions in areas of targeted growth' (Respondent to the online survey).

### **The writing test and alignment to the Australian Curriculum**

NAPLAN tests, including the writing tests, preceded the Australian Curriculum: English. To date, there has not been a published mapping of the English curriculum, including the Achievement Standards, the General Capabilities, the National Literacy and Numeracy Learning Progressions, the NAPLAN writing test and the Marking Guide. In its current form the writing test aligns with the English curriculum through its inclusion of the text types, imaginative and persuasive to date, and its focus on seven sub-strand threads of the curriculum – purpose, audience and structures of different types of texts; vocabulary; text cohesion; sentences and clause level grammar; word level grammar; punctuation; and spelling.

The writing skills in Foundation to stage 2, stages 3 to 4 and 5 to 6 are clear in the curriculum. The stated intent of the curriculum in secondary school (Years 7 to 10) is that learning in English 'builds on concepts, skills and processes developed in earlier years, and teachers will revisit and strengthen these as needed' (ACARA, 2019a, unpaginated).

ACARA (2018c) has drawn useful comparisons between the Australian Curriculum: English and the British Columbia New Curriculum English Language Arts (BCC:ELA), with both curricula built on the understanding of students becoming reasonably independent writers by Year 10.

By Year 10, Australian students are expected to be able to construct sustained texts for a range of purposes that address challenging and complex issues. Their writing should reflect an emerging sense of personal style, use of appropriate structure and use of language and literary devices and features which have been selected specifically for the intended audience.

The BCC: ELLA Composition Year 10 course develops students' skills in written communication. The course requires students to explore and create coherent, purposeful compositions through processes of drafting, reflecting and revising to create texts that demonstrate breadth, depth and evidence of writing for a range of situations.

Both curricula are built on the implicit understanding that students have, by now, become reasonably independent writers. Instruction is centred on writing techniques that allow students to craft and refine their writing for very particular purposes. For mastery of the content in either curriculum, students must be proficient in the fundamentals of writing, be able to plan, draft and edit, be skilled in accessing and applying research material and be able to select and use language forms and features in precise and accurate ways (ACARA, 2018c, p. 57).

The Australian Curriculum: English and the Singapore Curriculum: English Language Syllabus have in common that they “are built on the expectation that by the conclusion of compulsory schooling, students are independent writers with control over essential grammar, spelling and punctuation.” (ACARA, 2018d, p.57). Further, writing techniques are foregrounded in both curricula:

Both curricula are centred on writing techniques that allow students to craft and refine their writing for very particular purposes. By Year 10, for example, Australian students should be able to construct sustained texts that address challenging and complex issues. Their writing should reflect an emerging personal style, use of appropriate structure and the deliberate choice of language and literary devices to suit the purpose (ACARA, 2018d, p. 64).

The above segments make clear an expectation drawn from international practice that by Year 10, students will be proficient in strategies and skills of writing - what is referred to in BCC:ELA as ‘the fundamentals of writing’ (ACARA, 2018c, p.57) - and be able to select and use language forms and features with well-developed control of composing processes. This expectation needs to be considered against a backdrop of the NAPLAN writing data.

## Sex

The percentages of male and female students from 2008 to 2019 whose writing was judged to be below the National Minimum Standard (NMS) are shown in Table 19 (ACARA, 2020j). As previously mentioned, the below NMS figures stated in this chapter exclude exempt students. In 2019, nationally the rates for Year 3 were 2.8% for male students and 1.2% for female students. By Year 9, the rates were 21.3% for male students and 10% for female students. The increasing percentages of students with writing assessed as below NMS across NAPLAN testing year levels provide an opening to consider whether students have become ‘reasonably independent writers’ by Year 10. The information presented in Table 19 has been compiled using data from the NAPLAN National Reports (ACARA, 2020j), with a focus on NAPLAN writing results for males and females.

**Table 19:** Percentage of male and female students below National Minimum Standard in writing

Genre (below NMS)	Year	Sex	Year 3 (band 1)	Year 5 (band 3 and below)	Year 7 (band 4 and below)	Year 9 (band 5 and below)
<b>Narrative</b>	2019	Male	2.8	7.8	12.8	21.3
		Female	1.2	3.3	5.1	10.0
<b>Persuasive</b>	2018	Male	5.4	11.6	15.9	24.2
		Female	2.2	5.1	6.8	12.6
<b>Persuasive</b>	2017	Male	3.7	9.0	14.5	22.2
		Female	1.3	3.7	5.9	10.4
<b>Narrative</b>	2016	Male	2.7	7.2	12.1	20.8
		Female	0.9	2.7	4.8	9.4
<b>Persuasive</b>	2015	Male	3.7	8.2	15.3	23.8
		Female	1.6	3.4	6.5	11.3
<b>Persuasive</b>	2014	Male	5.8	10.7	13.7	22.5
		Female	2.5	4.5	5.5	10.2
<b>Persuasive</b>	2013	Male	4.4	9.2	13.1	22.1
		Female	1.7	3.4	4.8	9.0
<b>Persuasive</b>	2012	Male	3.8	8.3	12.3	23.1
		Female	1.5	3.3	4.5	10.1
<b>Persuasive</b>	2011	Male	3.9	8.0	10.5	19.1
		Female	1.6	3.2	3.9	8.0
<b>Narrative</b>	2010	Male	3.3	7.2	8.6	16.1
		Female	1.3	2.8	2.9	6.2
<b>Narrative</b>	2009	Male	3.4	7.4	9.0	15.6
		Female	1.4	3.1	3.3	6.0
<b>Narrative</b>	2008	Male	4.0	8.3	10.0	16.3
		Female	1.7	3.4	3.8	6.6

Note: The figures above refer to the lowest band and exclude exempt students.

**Table 20:** Percentage of students by location below National Minimum Standard in writing

		Year 3	Year 5	Year 7	Year 9
<b>NSW</b>	Major cities	0.9	3.5	6.6	12.4
	Inner regional	1.8	7.6	12.9	21.4
	Outer regional	2.5	10.9	16.9	29.6
	Remote	4.5	10.1	26.2	44.7
	Very remote	6.9	26.2	35.8	51.2
<b>Vic</b>	Major cities	1.1	1.9	4.6	10.7
	Inner regional	1.6	3.8	8.1	15.4
	Outer regional	1.6	3.7	9.7	15.5
	Remote	0.8	0.5	4.6	6.4
	Very remote	-	-	-	-
<b>Qld</b>	Major cities	2.0	6.0	9.4	18.1
	Inner regional	2.7	8.6	14.9	24.4
	Outer regional	3.4	10.4	16.9	27.1
	Remote	8.3	17.6	27.4	40.1
	Very remote	17.2	36.2	46.8	58.6
<b>WA</b>	Major cities	1.5	4.4	7.2	10.7
	Inner regional	2.4	7.9	11.1	15.7
	Outer regional	4.1	10.3	15.9	18.4
	Remote	7.0	15.5	20.5	26.8
	Very remote	21.1	36	48.7	57
<b>SA</b>	Major cities	2.0	6.5	7.4	13.6
	Inner regional	3.0	9.0	9.9	17.6
	Outer regional	3.8	12.1	13.8	25.6
	Remote	4.3	9.5	15.2	21.8
	Very remote	28.8	40.7	46.2	42.4
<b>Tas</b>	Major cities	-	-	-	-
	Inner regional	2.4	6.9	10.5	16.8
	Outer regional	3.0	9.8	14.4	22.0
	Remote	6.3	14.3	15.9	33.0
	Very remote	n.p.	n.p.	n.p.	n.p.
<b>ACT</b>	Major cities	1.9	4.2	8.2	13.5
	Inner regional	n.p.	-	-	-
	Outer regional	-	-	-	-
	Remote	-	-	-	-
	Very remote	-	-	-	-

		Year 3	Year 5	Year 7	Year 9
<b>NT</b>	Major cities	-	-	-	-
	Inner regional	-	-	-	-
	Outer regional	6.2	13.5	21.1	32.8
	Remote	21.0	32.1	42.3	46.4
	Very remote	67.9	79.3	88.2	90.9
<b>Aust</b>	Major cities	1.3	3.9	6.8	13.0
	Inner regional	2.1	6.8	11.6	19.7
	Outer regional	3.3	10	15.8	25.1
	Remote	9.0	17.2	25.0	32.8
	Very remote	33.8	49.1	61.2	68.2

Key: '-' or missing row indicates that the geolocation code does not apply within this state/territory or for this year level.

'n.p.' indicates data not published as there were no students tested or the number of students tested was less than 30.

## Geographic location

In addition to the marked differences among male and female students and across the years of schooling in the percentages of students below NMS, there are also marked geographic differences as shown in Table 20, where schools are classified using the Australian Bureau of Statistics' Australian Statistical Geography Standard Remoteness Structure with Major Cities of Australia, Inner Regional Australia, Outer Regional Australia, Remote Australia and Very Remote Australia (ACARA, 2020j).

The data show a concerning picture of writing performance in remote and very remote locations in particular. In three states and one territory (NSW, Queensland, Western Australia and the Northern Territory), the NAPLAN writing test results show more than 50% of students in Year 9 have been assessed as below the NMS.

At the national level, the data show a stark difference between the percentages of Year 9 students in major cities (13%) and those reported for remote (32.8%) and very remote (68.2%) locations. Being below NMS for Year 9 is defined as Band 5 or below on the NAPLAN scale.

The descriptions of the ten NAPLAN writing proficiency bands in (ACARA, 2020e) are shown in Table 21. The NAPLAN assessment scale is provided in Chapter 4 Figure 7.

**Table 21:** Descriptions of performance bands on the writing scale

Proficiency band	Writing skills and knowledge
<b>Band 10</b>	Writes a cohesive, engaging text that explores universal issues and influences the reader. Creates a complete, well-structured and well-sequenced text that effectively presents the writer's point of view. Effectively controls a variety of correct sentence structures. Uses punctuation correctly, including complex punctuation. Spells all words correctly, including many difficult and challenging words.
<b>Band 9</b>	Incorporates elaborated ideas that reflect a worldwide view of the topic. Makes consistently precise word choices that engage or persuade the reader and enhance the writer's point of view. Punctuates sentence beginnings and endings correctly and uses other complex punctuation correctly most of the time. Shows control and variety in paragraph construction to pace and direct the reader's attention.
<b>Band 8</b>	Writes a cohesive text that begins to engage or persuade the reader. Makes deliberate and appropriate word choices to create a rational or emotional response. Attempts to reveal attitudes and values and to develop a relationship with the reader. Constructs most complex sentences correctly. Spells most words, including many difficult words, correctly.
<b>Band 7</b>	Develops ideas through language choices and effective textual features. Joins and orders ideas using connecting words and maintains clear meaning throughout the text. Correctly spells most common words and some difficult words, including words with less common spelling patterns and silent letters.
<b>Band 6</b>	Organises a text using paragraphs with related ideas. Uses some effective text features and accurate words or groups of words when developing ideas. Punctuates nearly all sentences correctly with capitals, full stops, exclamation marks and question marks. Correctly uses more complex punctuation markers some of the time.
<b>Band 5</b>	Structures a text with a beginning, complication and resolution, or with an introduction, body and conclusion. Includes enough supporting detail for the text to be easily understood by the reader, although the conclusion or resolution may be weak or simple. Correctly structures most simple and compound sentences and some complex sentences.
<b>Band 4</b>	Writes a text in which characters or setting are briefly described, or in which ideas on topics are briefly elaborated. Correctly punctuates some sentences with both capital letters and full stops. May demonstrate correct use of capitals for names and some other punctuation. Correctly spells most common words.

Proficiency band	Writing skills and knowledge
<b>Band 3</b>	Attempts to write a text containing a few related events or ideas on topics, although these are usually not elaborated. Correctly orders the words in most simple sentences. May experiment with using compound and complex sentences but with little success. Orders and joins ideas using a few connecting words but the links are not always clear or correct.
<b>Band 2</b>	Shows audience awareness by using common text elements, for example, begins writing with Once upon a time; or I think ... because ... Uses some capital letters and full stops correctly. Correctly spells most simple words used in the writing.
<b>Band 1</b>	Writes a small amount of simple content that can be read. May name characters or a setting; or write a few content words on a topic. May write some simple sentences with correct word order but full stops and capital letters are usually missing or incorrect. Correctly spells a few simple words used in the writing.

Referring to data in Table 19 and Table 20, the implicit understanding that students have become reasonably independent in writing by Year 10 appears problematic. At a deeper level, if writing is understood to be a key means through which students learn in all curriculum areas, then the data point to how Year 9 students whose writing is assessed as below NMS are likely to face significant barriers to success in senior schooling, given that they are not able to write in ways described in bands 6 to 10.

### High performance in writing at Year 9

The percentages of Year 9 students who were in the top two bands (bands 9 and 10), and so well above National Minimum Standard, are shown in Table 22 (ACARA, 2020j). This table indicates there are generally declining percentages of students in the top 2 bands from 2011 to 2019. At the national level in 2011 there were 21.5% at this level (13.4% achieving at band 9 and 8.1% achieving at band 10). In 2019, only 12.4% of students achieved at this level (9.4% at band 9 and 3% at band 10).

**Table 22:** Percentages of Year 9 students in top two bands in NAPLAN writing

	2011	2012	2013	2014	2015	2016	2017	2018	2019
	Per	Per	Per	Per	Per	Nar	Per	Per	Nar
<b>NSW</b>	19.8	18.8	17.3	15.2	13.4	11.7	16.8	14.1	13.1
<b>Vic</b>	25.4	19.4	18.2	15.7	15.3	15.5	16.6	11.4	12.8
<b>Qld</b>	20.3	11.2	14.2	12.6	11.3	8.5	12.6	8.6	9.2
<b>WA</b>	20.6	18.3	16.8	17.3	14.5	13.3	16.3	13.9	15.6
<b>SA</b>	20.5	14.9	15.5	14.1	12.9	10.9	12.1	9.4	13.8
<b>Tas</b>	17.3	14.2	13.4	11.4	10.4	12.8	13.0	7.8	12.8
<b>ACT</b>	26.3	20.7	21.8	18.2	17.0	13.7	19.5	15.0	14.4
<b>NT</b>	14.7	9.8	9.2	9.4	6.1	7.7	10.5	8.3	6.4
<b>Aus</b>	21.5	16.8	16.5	14.8	13.4	12.3	15.4	11.7	12.4

Key: Per=Persuasive genre; Nar=Narrative genre

The writing results presented in Table 19, Table 20 and Table 22 provide a basis for considering greater explicitness about the writing knowledge and skills that students are expected to develop in the upper middle years. This revisits the implicit understanding made in the current Australian Curriculum: English that students become reasonably independent writers by Year 10. The data suggest that this could be considered aspirational. The data also open the space to consider a strengthened focus on the writing domain in the current review of the Australian Curriculum: English and other curriculum areas, and how these are intended to align with the National Literacy and Numeracy Learning Progressions.

### Critique of the writing test

There are 10 specified criteria in the rubric or scoring traits used for assessing writing in NAPLAN. Nine of these are common across the NAPLAN testing year levels, with some customisation occurring in criterion 4, dependent on the selected form of writing (persuasive or narrative writing). The criteria

are not weighted equally. The suite of criteria and the related score ranges are:

1. Audience [0-6]
2. Text structure [0-4]
3. Ideas [0-5]
4. Character and setting [0-4]  
(for narrative writing)  
Persuasive devices [0-4]  
(for persuasive writing)
5. Vocabulary [0-5]
6. Cohesion [0-4]
7. Paragraphing [0-3]
8. Sentence structure [0-6]
9. Punctuation [0-5]
10. Spelling [0-6].

As shown, the 10 criteria are to be marked on four different score ranges [0-3, 0-4, 0-5, 0-6], and then totalled to compute a composite score from an available total of 48.

Paragraphing has the most limited range; text structure, persuasive devices, character



and setting, and cohesion are slightly higher, with a range of up to 4 possible points; ideas, vocabulary and punctuation, up to 5 points; and audience, sentence structure and spelling each with a range up to 6 points. The consultations brought forward considerable dissatisfaction, not only with the criteria, but also with the accompanying scores. They were talked about as having a distorting effect on teaching writing.

More than two decades ago, Messick (1994) posed the issue of whether rubrics validly meet the purposes of their usage and the prospects of their achieving valid assessments of performance, especially in the case of complex performances. His question was:

‘By what evidence can we be assured that the scoring criteria and rubrics used in holistic, primary trait, or analytic scoring of products or performances capture the fully functioning complex skill?’ (p. 20).

This question has relevance to this review. Only limited empirical research has been conducted to date concerning the nature and function of rubrics in education generally, and the NAPLAN writing rubrics (narrative and persuasive) in particular.

Research by Humphry and Heldsinger (2014) is a notable example. They identified what they referred to as ‘a potentially widespread threat to the validity of rubric assessments that arose due to design features’ (p. 253) and presented evidence from empirical research conducted in the context of assessing narrative writing using rubrics. They claimed that:

the evidence indicates that the typical grid or matrix design of the rubric design used in this context [narrative writing] induces pronounced rating tendencies of a form that would usually be interpreted to indicate a halo effect. The term halo effect refers to a strong tendency for

ratings on separate items or criteria to reflect a general rater impression of a performance (pp. 253-254).

In this empirical investigation of two different rubrics,

It was established that the issue was not the raters’ inability to treat each criterion independently but that the rubric itself forced judgements to be dependent, resulting in an apparent halo effect (p. 262).

While Humphry and Heldsinger (2014) did not argue that this finding was generalisable to rubrics in other contexts, they proposed the need to resolve the validity effect potentially caused by the design of scoring rubrics. They also asserted the value of:

more productive research into a number of questions, such as to ascertain which and how many criteria should be used, whether the operational independence of criteria can be established, and the optimal number of qualitative gradations for each separate criterion. Resolving the threat to validity might also open the way to more productive research into whether raters make more valid assessments using rubrics than holistic judgments (p. 253).

Regarding the number of criteria, other researchers (for example, Sadler, 1989) have suggested that fewer criteria may be desirable to achieve high rater-consistency (with self and over time) and inter-rater reliability (self with other raters). In NAPLAN writing it is arguably unreasonable to expect individual markers to hold 10 criteria in their head concurrently during a scoring episode. ACARA indicated its intent to examine the reliability and validity of the scoring rubric developed to incorporate these criteria of the writing task as part of its next four-year work plan, with revisions to be made as warranted (ACARA, 2017). To date, however, there have been no published revisions to

the criteria. Calls for the revisions echoed through the consultations with one system/sector, one of many respondents making the case, saying that it 'also supports a review of the marking rubrics as the current criteria do not reflect the reality of teaching writing in Australian classrooms'.

### **Conventions of language test**

The ACARA Assessment Framework: NAPLAN Online (2017-2018) states that the 'NAPLAN writing test complements the NAPLAN conventions of language test assessing spelling, grammar and punctuation within the context of writing' (p. 14). The additional language conventions test includes three categories namely the grammar and punctuation items, and the spelling test.

The grammar items in the grammar and punctuation test focus on knowledge and accurate use of grammar at a sentence, clause and word level. Grammar items are developed from the content of the Australian Curriculum: English sub-strand threads of text cohesion, sentences and clause level grammar and word level grammar. (ACARA, 2017, p. 12)

The punctuation items in the test focus on the identification of accurate use of punctuation conventions. Punctuation items are developed from the content of the Australian Curriculum: English sub-strand thread of punctuation. (ACARA, 2017, p. 12)

The NAPLAN spelling test focuses on the accurate spelling of written words, and consists of an audio component and a proofreading component. Spelling items are developed from the Australian Curriculum: English sub-strand thread of spelling. (ACARA, 2017, p. 13)

In presenting the NAPLAN Online language conventions test an innovation includes the use of an audio file 'in which words

are presented in context sentences' (ACARA, 2017, p.13) with accommodations to be made for students with hearing impairments. A further innovation is the interlocking grammar and punctuation testlet design as part of the adaptive design in NAPLAN Online. Respondents indicated that the inclusion of spelling and punctuation in both the writing test and the language conventions test appeared to be unnecessary, especially if the writing test were to be 'improved', as indicated below. This observation applied irrespective of the mode of the NAPLAN writing test. A preference for assessing language conventions as part of the writing test was repeatedly mentioned.

Do we need to continue to assess language conventions in the NAPLAN test? If we improve the writing test, we could incorporate language conventions as part of writing (Subject association).

The available scores for spelling, punctuation, paragraphing, and grammar were reported to be at the expense of higher-order writing features. The NAPLAN narrative and persuasive marking manual (ACARA, 2010, 2013a) indicates that scorers are to count the occurrence of correctly spelt words defined as simple, common, difficult and challenging. A script containing no conventional spelling scores zero, with correct spelling of most simple words and some common words yielding a mark of two. To attain a mark of six, a student must spell all words correctly, and include at least 10 difficult words and some challenging words or at least 15 difficult words (ACARA, 2010, 2013a). Respondents' widely held view, also reported by Perelman (2018), that this approach to scoring spelling had the effect of prioritising accurate spelling of easier words rather than attempted approximations of more difficult vocabulary that accrue minimal score – 'The writing test marking criteria encourage a formulaic

writing style. The marking criteria privilege grammar and spelling over ideas' (Subject association).

It was also reported that NAPLAN writing encourages a response using a five-paragraph form. According to Perelman (2018), 'Although the five-paragraph essay is a useful form for emerging writers, it is extremely restrictive and formulaic. Most arguments do not have three and only three supporting assertions. More mature writers such as those in Year 7 and Year 9 should be encouraged to break out of this form. The only real advantage of requiring the five-paragraph essay form for large-scale testing appears to be that it helps ensure rapid marking' (p. 8).

There was a large corpus of commentary on the current criteria for the NAPLAN writing test, with the clear thread being the need for a review of the criteria, the related numeric scoring, the choice of prompts and the limited range of forms set in the test.

The writing component of NAPLAN is problematic for a number of reasons including the prevalence of rehearsed or formulaic writing; the accuracy of assessment rubrics and criteria and the challenge of effectively assessing in online environments. ... While acknowledging the concerns... [it] remains supportive of the writing component continuing at the present time. (School system/sector)

[It] welcomes the review of the content suite covered by NAPLAN testing. An evaluation of the writing task is especially appreciated, though [it] questions the value of removing the writing task. (School system/sector)

### Markers and scoring

NAPLAN marker training is undertaken within states and territories, with oversight by ACARA. Teachers and other suitably qualified personnel in each jurisdiction

are invited to apply to be NAPLAN markers. Sample scripts are provided to all markers nationwide; state-based marker quality teams use these for locally implemented training and calibration. There is little published information regarding processes for achieving training consistency nationwide and monitoring of local processes to maintain marking quality within jurisdictions over time. Limited information is available, including in technical reports, regarding how marker feedback processes are managed during scoring operations at jurisdictional and national levels.

The scoring process involves a single marker who scores each student's script, applying all 10 criteria. A possible effect of this is that the correlation between scores on different criteria could be artificially inflated, leading to possible violations of the local independence assumption built into the scaling model (Humphrey & Heldsinger, 2014). A study to examine such effects should be undertaken given the concerns expressed by a number of stakeholders. Further options to explore in marking implementation include: i) multiple markers marking a subgroup of students' scripts and ii) different markers assigning marks to different criteria for a particular student – 'There are validity issues where teachers mark writing tests too fast and to set criteria. The criteria and allocation of marks should be closely examined' (Subject association).

Teachers' professional judgement was repeatedly referred to as a critical missing element in NAPLAN writing test processes. The recurring and strongly expressed view was that the re-visioning of NAPLAN writing should make explicit provision for teacher judgement, including a component of moderation to seek comparability across teachers and schools.

These observations open a space to consider the potential forms of accountability and verification of systems and processes that could be nationwide. Addressing these matters could build confidence in reporting. One respondent pointed to the need for bolstering confidence in the test stating,

‘There is not a lot of confidence in the writing test. Those running the assessment have indicated that writing is the area of NAPLAN that they are least confident in for reliability’  
(School system/sector).

### **NAPLAN writing test mode**

Test mode (computer-based writing test and a pen and paper test) featured as an important issue throughout the stakeholder consultation process. The widely reported view was that the teaching of handwriting should be prioritised over the teaching of keyboarding in the early years, and that the pen and paper test mode was frequently mentioned as ‘the only’ defensible choice for students in Year 3. A widely held view was that Year 3 students were too young to sit the writing test online with one respondent stating, ‘they should be properly focusing on handwriting and learning how to write’.

This stance is consistent with the findings from the Centre for Education Statistics and Evaluation (CESE)’s 2016 study that investigated,

‘whether primary students in NSW schools perform differently according to the mode of the writing test... and the extent to which typing proficiency accounts for any differences observed in students’ performance in a computer-based writing test versus in a pen and paper test’ (Lu, Turnbull, Wan, Rickard & Hamilton, 2017, p. 4).

The study found that, for Year 3 students, ‘the median typing speed was 9 words per minute’, which, based on the literature, is

reported to be ‘lower than the handwriting speed for this age group, hence it is likely that many Year 3 students would struggle to produce online texts comparable to handwritten texts in a timed condition’ (p. 5).

A related key finding is that ‘most trial schools do not explicitly teach keyboarding skills’ (p. 4). Consistent with other published research, the recommendation is that ‘typing instruction is best commenced in the upper primary years’ (Lu et al., 2017, p. 5). Perhaps more importantly, the study recognised i) the need to investigate how new technologies can be used to enrich the teaching of writing and students’ experience of the writing process, and ii) for schools to identify ‘an effective method for developing students’ typing fluency and to monitor the development of their typing proficiency over time, for students beyond Year 3’ (p. 5).

During the review consultations, participants identified how keyboarding had potential to add to the cognitive load of students in sitting the writing test online. There were also numerous comments regarding the technical readiness of schools to participate in NAPLAN Online, an observation also reported in the 2016 CESE study. Some review participants mentioned the limited number of computers available in some schools, the limited technical support available, and the challenges faced by some schools relating to system infrastructure and school budgets. These issues go beyond this review but are recognised as core in enabling all students to have equity of opportunity for success in participating in the writing test online, and NAPLAN Online more generally. The panel is aware that systems/sectors are putting in change programs to help schools transition to online testing.

It was clear that many schools welcomed the move to NAPLAN writing online, indicating ‘we were ready for this’. However, there were recurring concerns about the

preparedness of some students in Years 5, 7 and 9 'to do their best writing' in time-restricted conditions, especially for those with limited typing fluency.

Several respondents identified that those students in Years 5, 7 and 9 who had prior opportunities to develop typing fluency were better placed to produce sustained writing online, while those students with little, or no, experience, were characterised as being at a disadvantage. The explicit teaching of typing fluency in the curriculum could go some way to address this.

The move to NAPLAN writing online was consistently referred to as having significant resourcing implications for preparing students for sitting the test in schools. Operational impediments including those relating to school and system/sector infrastructure were also reported.

Concerns regarding NAPLAN Online Automated Writing Evaluation were widespread. This was the case, even though respondents recognised that this would lead to earlier return of writing test results to the schools and that, for many students, their 'normal' is working online.

While there was a full range of views about the move of NAPLAN to online, in the case of the writing test, by far the dominant view was that student writing should not be machine marked: 'marked by a robot'. While many recognised that 'a machine' can score technical or grammatical criteria (for example, sentence structure, punctuation and spelling), there was explicit rejection of the idea that genre-based or authorial criteria (for example, audience, ideas and cohesion) of writing could be fairly assessed using Automated Writing Evaluation. More comprehensive research could address concerns about the utility and efficiency of the machine scoring technical and authorial criteria.

The NAPLAN Technical Report (ACARA, 2020e) examined jurisdictional Differential Item Functioning for both paper and online marking for the writing test, finding that "the expected score curves of the ten ratings criteria were plotted for the eight jurisdictions...None of the criteria showed notable differences across jurisdictions" (p. 100).

There is a need to examine a range of strategies for building confidence in the profession and in the community regarding the reliability of Automated Writing Evaluation scoring and its dependability in assessing all aspects of writing. Implications of Automated Writing Evaluation for teacher professionalism merit serious consideration. Further research designed to investigate and collect evidence on the feasibility and validity of automated scoring systems and processes will be essential in the context of NAPLAN writing. These next steps would build on already completed work by ACARA, and ongoing international research and development, to generate necessary empirical evidence on the validity of automated scoring systems in relation to the validity of tests content and response processes including scoring (ACARA, 2018f; Bridgeman & Ramineni, 2017; Shermis 2014; Eliot et.al, 2013).

The data that the writing test has produced is distinctive internationally, and its potential for longitudinal investigations of children's and young people's trajectories in writing is arguably under-utilised. Longitudinal research would be greatly facilitated by the availability of a Unique Student Identifier with which to link students' writing in future tests.

## Summary

The calls for change presented in this section of the review go well beyond adding another form (for example, informative writing). They

give voice to a wide range of significant issues with the NAPLAN writing test.

These include:

- the content of the writing tests, including the prompts, choice of forms and the criteria and related range of scores
- technical issues of scoring including dependencies among criteria, especially in relation to adjacent year levels. This opens the opportunity for investigating the difficulty and ease with which scorers can separate the criteria for scoring purposes, discriminating among them and addressing the specified features within each criterion
- the conditions under which students sit the test, providing limited opportunity for planning, drafting, revising and editing
- the validity and reliability of the marking operations currently undertaken within jurisdictions and the potential for a strengthened form of monitoring that could be nation-wide.

The data suggest that the writing test in its current form has not sustained confidence and trust within the profession and at system level. Writing results are perceived to be less reliable than those from the reading and numeracy tests. Further, writing does not feature strongly in improvement targets and strategies in some states and at school level. Overall, it is clear that there is little, if any, support for a claim that the test is perceived to be well-aligned with classroom writing and further, that it has positively impacted on the teaching of writing, and that students have benefited from participation – ‘The singular focus on persuasive writing has not had a good impact on teaching and student writing is formulaic’. (Education expert)

The calls for significant change to the writing test were clear and sustained throughout

the consultations with teachers, school leaders, other experts in systems/sectors and unions, and professional associations.

Calls for change should be distinguished from a uniform call for removing standardised assessment of writing as a domain in NAPLAN. The preceding discussion shows that, while several areas of dissatisfaction and concern were identified, there remains clear support for standardised assessment of writing to continue in NAPLAN, following significant redevelopment. The need for a redeveloped writing test was the dominant position. Regarding next steps, however, there is support for national testing of writing through sampling, and also as a census test. Chapter 7 takes up this matter in the recommendations.

Finally, the review brought to light strong support for rethinking the role of the profession in national testing including for system monitoring purposes, and the role of teacher judgement in scoring and interpreting NAPLAN writing results in conjunction with other assessments to inform practice. It is also clear that further research is needed into Automated Writing Evaluation in the context of NAPLAN writing. This review has opened the opportunity to consider the potential to bring together human judgement, moderation and machine scoring: Automated Writing Evaluation scoring of language conventions (for example, grammar, punctuation, spelling<sup>5</sup>, paragraph, sentence structure) with authorial aspects of writing (for example, audience, text structure, ideas, character/setting/persuasive devices, vocabulary, cohesion) scored by teachers.

Chapter 7 takes up the issues in NAPLAN writing to be addressed and presents recommendations for action.

---

5 Grammar, punctuation and spelling are currently tested in the NAPLAN language conventions test.

# Chapter 6. Uses of NAPLAN

Chapter 1 identified five purposes for the current national standardised assessment program – monitoring progress towards national goals, school system accountability and performance, school improvement, individual student learning achievement and growth, and information for parents/carers on school and student performance. The purpose of this chapter is to describe how NAPLAN data are used nationally, by school systems/sectors and schools, by teachers and by parents/carers and the broader community.

## Key points

- For more than a decade, NAPLAN's standardised assessment data have underpinned public reporting on national, state and territory trends in student achievement and growth.
- Wide differences in school-level achievement and growth may be observed among schools serving similar communities, and these differences are reported to schools through systems/sectors' data analytics tools and to the public through My School and school system/sector websites.
- States and territories and school systems/sectors set targets for NAPLAN achievement, often in terms of increasing the proportion of students in higher achievement bands and decreasing the proportion in lower bands.
- School-level NAPLAN results are routinely used by schools in target-setting, planning and monitoring achievement, but some stakeholders believe that school-level NAPLAN targets narrow teachers' focus to students near the boundaries of bands of achievement and unfairly categorise schools with lower than expected student achievement.
- Teachers report that they use individual-level NAPLAN results in triangulation with other standardised assessments and teachers' judgments.
- Parents/carers value individual-level NAPLAN results as a source of information external to students' schools.
- Many professional stakeholders are opposed to the publication of school-level NAPLAN results because they can be used for school comparisons and league tables, but there is a tension between this view and evidence of broader community expectations about transparency of school-level achievement data.

## National uses

The annual National Reports published each year since 2009 have used whole-cohort NAPLAN assessments to report on trends in achievement, alongside a range of other educational indicators including attendance, participation and school funding. These reports also provide detailed analysis of achievement and cohort gain in each of the NAPLAN cohort assessment domains at national and jurisdictional levels, with subsidiary analysis of achievement by sex, Indigenous status, language background other than English status, geolocation and parental education.

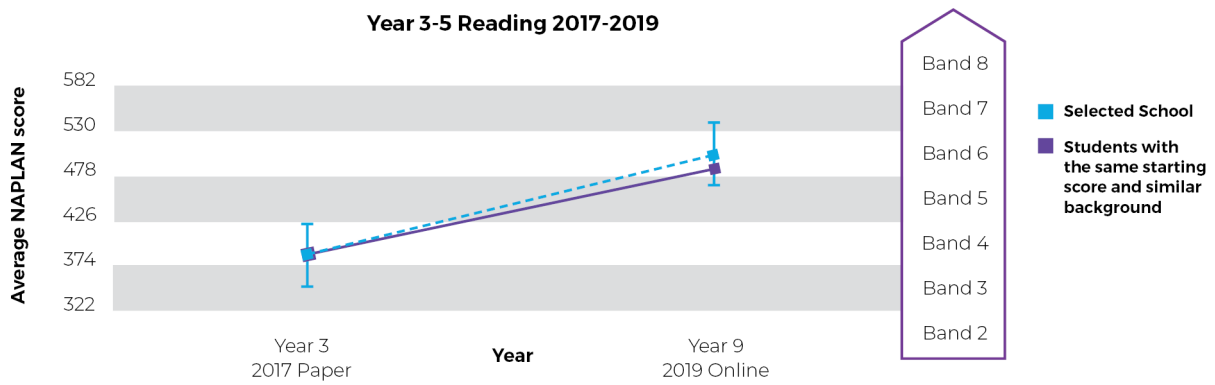
The use of national standardised assessments to monitor national, state or territory trends or to monitor the performance of equity groups was not contested by stakeholders consulted for the NAPLAN Review. School sector respondents noted that the data were used to identify areas of need, including equity issues. The national interests most often mentioned by online respondents included government decision making and funding allocation to areas of the greatest need, consistent with the support among stakeholders for the national monitoring purpose of NAPLAN identified in Chapter 1. As we have documented elsewhere, however, concerns about the use of NAPLAN for school comparisons and league tables were widespread and many stakeholders suggested that the national interest in monitoring achievement could be met equally well with sample assessments.

Because the kinds of simple comparisons of average achievement that often appear in league tables are misleading, governments and school systems/sectors have looked towards more nuanced comparisons that take account of differences in the backgrounds and achievement of students served by each school. The first version

of the My School website avoided league tables but did provide each school with a comparison with other schools with students from similar home backgrounds. In framing the first version of My School, data on individual families in schools were not available so the Australian Curriculum, Assessment and Reporting Authority (ACARA) used data from Australian Bureau of Statistics Collection Districts in which students lived. Using collection district data, ACARA created a new measure of advantage based on the education and occupations (but not income) of adults in the district. It was labelled an Index of Community Socio-educational Advantage (ICSEA) and not as an index of 'socio-economic advantage'. The technical weakness of this strategy is that it makes the statistical assumption that the social characteristics of a district could be applied to all students regardless of the school they attended. For the 2011 version of the My School website, data on individual families were available so it was no longer necessary to use collection district data. ACARA developed a new measure of advantage based on the actual parents/carers' education and occupation. Now that it used family data and not community data, it might have been labelled an Index of Socio-educational Advantage (ISEA) to signal the change but the established name and the acronym ICSEA were retained.

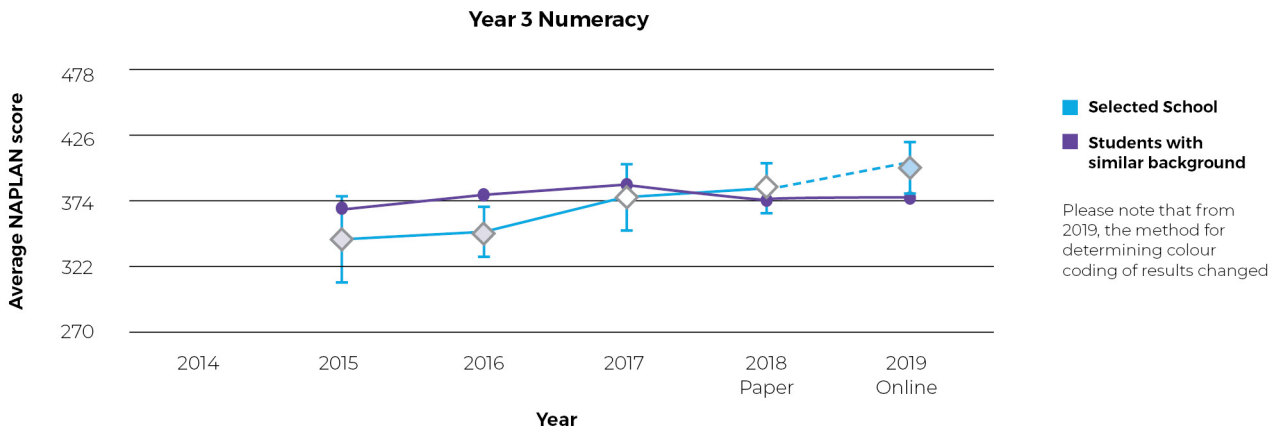
Following the NAPLAN Reporting Review (Louden, 2019), education ministers agreed to remove the 'similar schools' display from the 2019 version of My School and to introduce new measures of student progress. Instead of comparisons among 'similar schools', one school-level display compares improvement with that of other students across the country who had the same NAPLAN score two years ago and who have a similar background as the students at the selected school (Figure 8).





**Figure 8:** My School comparison with students with same starting score and similar background

A second display shows average achievement over time compared with students with a similar background (Figure 9).



**Figure 9:** My School: Selected school compared with students with and similar background

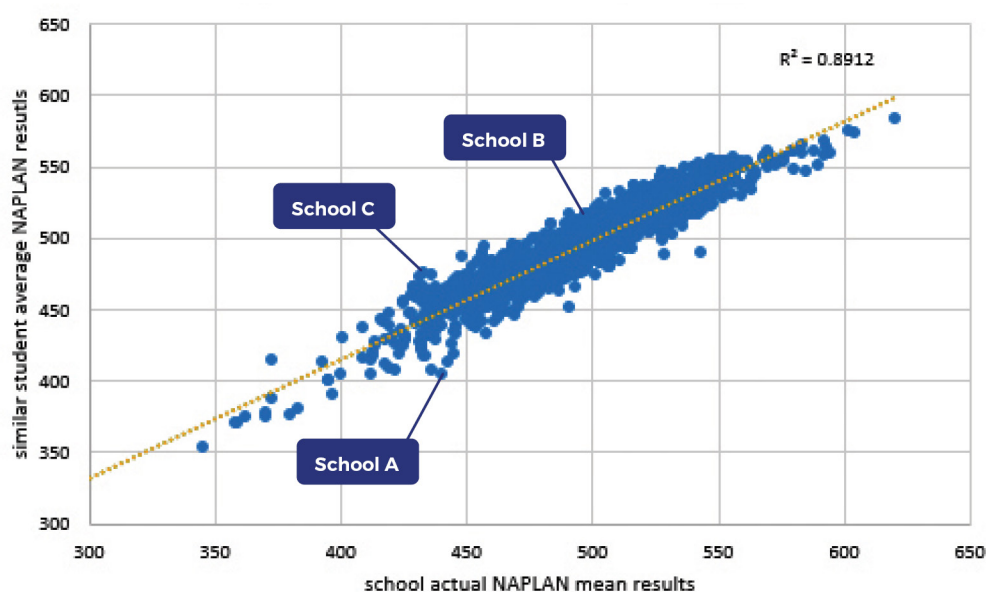
Notwithstanding the introduction of the 'same starting score' component and the elimination of the 'similar school' display, some stakeholders continued to express concern about whether ICSEA does effectively identify schools in comparable circumstances.

The use of the Index of Community Socio-Educational Advantage (ICSEA) calculations ... to make comparisons with schools is constructive but needs to be understood in a context where schools have multifaceted environmental factors or unique philosophical approaches. (Written submission response: School system/sector)

While the data on My School purports to be a means of comparing 'like schools' with 'like schools', the Index of Community Socio-Educational Advantage (ICSEA) measure does not always enable proper comparisons to be made. ICSEA is inadequate in that it fails to consider a range of additional variables about a student's background that may impact on NAPLAN performance. The use of ICSEA should be reviewed. (Written submission response: School system/sector)

More broadly, many stakeholders contributing to the review argued that the public availability of comparative data across schools created unreasonable stress on teachers and, in turn in some cases, on students. Some wanted to see the data no longer published. Others, expecting that available data could not be suppressed after nine years of publication since the first My School in 2010, proposed that data be collected only for a sample of students rather than for all.

The alternative view is that school-level data are already collected, that these data reveal significant differences in results achieved by schools serving similar student populations, and that these differences are a matter of legitimate public interest. Figure 10 provides an example using NSW Government school data from the 2019 ACARA My Schools data set. It shows the distribution of school actual average NAPLAN scores compared with average scores achieved by students with similar backgrounds. The diagonal yellow line is the regression line summarising the relationship. It slopes upwards, left to right, showing that, on average, schools with students from more advantaged backgrounds achieve higher average NAPLAN results. What is also clear is that many schools are not well-described by this average relationship. School C performs much better than would be expected given the level of advantage of its students. So does School B. School A, on the other hand, performs worse than would be expected given the level of advantage of its students.



**Figure 10:** Relationship between schools' socio-educational advantage and NAPLAN results

The data displayed in Figure 10 make clear that there are marked differences in NAPLAN results among schools serving students with similar backgrounds. School A could not readily substantiate a claim that it could not do better, 'given the kind of students it has'. Schools with advantaged students below the regression line to the right of the display may be comfortable with performances above the national average but they too are doing less well than other schools with similarly advantaged students. This may raise the question of whether these schools are 'coasting' and, like School A, can be challenged by comparison with other schools at a similar level of socio-educational advantage to improve their performance.

## School systems/sectors

The state and territory education systems/sectors use analyses like the one in Figure 10 to identify schools from which more could be expected and some were doing so with state and territory assessments before common national assessments were introduced with NAPLAN and before My School provided all schools with comparative data.

School systems/sectors make widespread use of NAPLAN data – including statistically similar school data – in setting system/sector-wide achievement targets and managing their school improvement programs. The 2018 National School Reform Agreement committed state and territory governments to improvements in academic achievement expressed in terms of NAPLAN targets:

[To] lower the proportion of students in the bottom levels and increase the proportion of students in the top levels of performance (bottom two and top two bands) in the National Assessment Program–Literacy and Numeracy (NAPLAN) Literacy and Numeracy, of Years 3, 5, 7 and 9 (COAG, 2018, p. 7).

This national commitment has cascaded through to plans and targets published by jurisdictions and systems/sectors. In Victoria, for example, the 2018-19 Annual Report of the Department of Education and Training reported on the indicator 'Students meeting the expected standard in national and international literacy and numeracy assessments' using NAPLAN literacy and

numeracy results in the top two and bottom three bands to (Victoria Department of Education and Training, 2019, p.19). Similarly, performance measures for the NSW Department of Education's 2018-2022 strategic plan include 'Increased proportion of students in the top two NAPLAN bands for reading and numeracy'. The 2019-20 targets for Queensland include the proportion of students at or above the NAPLAN National Minimum Standard in reading, writing and numeracy (Queensland Department of Education, 2019, pp. 7-8). The ACT's targets are set in terms of NAPLAN reading and numeracy gain scores Years 3 to 5 and 7 to 9. (Australian Capital Territory Budget Statements, 2019, pp. 7-8).

School systems/sectors pursue such state-level targets through their various school-level targets and improvement programs. To take just two examples, in Victoria the Differentiated School Performance Method has been developed to target support for schools. Data used in allocating schools to performance groups include levels of achievement in the top two and bottom two NAPLAN bands in Years 5 and 9 as well as NAPLAN benchmark performance growth (Victoria Department of Education and Training, 2019, p. 3). In NSW public schools, a targeted program called Bump it Up provides every school with tailored targets for improving performance in reading, numeracy, wellbeing and attendance. The Department of Education has reported that two-thirds of the Bump it Up schools improved their share of students in the top two NAPLAN bands between 2015 and 2019.

Jurisdictions have also developed digital dashboards and data analytics software to assist in tracking achievement and targeting improvement. The NSW Government's Scout system, for example, is described as having been developed to "provide school and corporate staff with information about what's working well, and what can be improved." The NAPLAN component of Scout provides online, graphics-intensive

information on school performance, student performance and NAPLAN item-level performance. The school performance component of Scout includes displays showing NAPLAN scores over time, the performance of equity groups; number and percentage of students in achievement bands over time; the percentage of students in each band compared with a statistically similar school group and the whole state percentage of students in the top two bands in reading and numeracy; and student growth in scores and across bands. Scout is available to government and non-government schools in NSW and the ACT. Similar systems linking NAPLAN and other data to school improvement planning include Queensland's OneSchool and Victoria's Panorama.

There was little consensus among stakeholders about the appropriateness of school systems/sectors use of NAPLAN data. School system/sector representatives often emphasised the value of NAPLAN in managing a large-scale system. As one put it – 'We have 2,200 schools in our system and we need to be able to initiate conversations with schools about their performance. NAPLAN helps us to frame that conversation' (School system/sector).

Members of the NAPLAN Review's Practitioners' Reference Group characterised NAPLAN as 'the main form of accountability' for public schools and described how 'annual implementation plan goals and school strategic plan goals are linked to NAPLAN data'. Some found this helpful and acknowledged that although "NAPLAN data is used to set targets for students and informs teacher's work quite heavily... It is also encouraging for these schools to see improvement in NAPLAN results."

Others noted how system/sector targets in some jurisdictions led to categorisation of schools based on their students' NAPLAN performance. Such categorisations could depend on the performance of a small

number of students at the edge of particular performance bands. ‘We lost three kids in the top two bands in reading last year, so now we’re a ‘transform’ school, which was quite devastating’ (Member of the NAPLAN Review Practitioners’ Reference Group).

School performance measures can narrow teacher focus to individual students that can be moved into the top two bands. Schools can be asked to identify a number of students to move out of middle bands into top bands and to then choose individual students to provide with targeted support. This means other students miss out. (Member of the NAPLAN Review Practitioners’ Reference Group)

Union stakeholders described NAPLAN as ‘overused’ in accountability and argued that systems/sectors’ use of NAPLAN in setting school-level achievement targets had undermined teaching.

By making NAPLAN results into the high stakes indicator of student attainment and school quality, governments and education departments undermined the fundamentals of good teaching and learning in schools.” (Written submission response: Union)

## Schools

The most commonly mentioned use of NAPLAN data in schools was in triangulation with other standardised assessment data and teachers’ professional judgements. NAPLAN results were variously used to ‘look at trends on cohorts’ and to consider ‘the skills that the school may need to focus on’. One said, “We love the student and school summary report (SSSR), and if we could have that straight away, it would be more useful.” Members of the Review’s Practitioner Reference Group, however, were clear that they did not over-emphasise NAPLAN results in their school planning – ‘NAPLAN

data can be used as a check-in point to consider teacher judgement within a school compared to other schools’; ‘We have heaps of data already, and this is the least accurate data we have. The accuracy just isn’t there.’; ‘Data is always useful, but we have other ways of getting data. We use PAT testing. All Catholic schools use the same learning management system, and the data uploads easily’; ‘We pin things down to specific areas, for example, if spelling dips down in comparison to previous years [...] We’ve done interventions in areas of targeted growth.’

We look at the strengths that come out and areas of weakness. If specific students are out line, we don’t worry if it is a bad day, but if it identifies an area they are not good at we use that. We are sometimes surprised how high our kids score on reading. It is the only time we compare to other schools and we don’t necessarily put a huge emphasis on the comparison as NAPLAN doesn’t test the things we value like creativity, problem-solving or thought.

Stakeholders reported widespread use of NAPLAN data analytics tools. Catholic sector representatives mentioned the Catholic Education Network’s (CENet) CED3 portal as well as a range of locally developed tools and reports. Others mentioned public school systems/sectors’ data analytics tools and the raw data files available from curriculum and assessment authorities. Despite teachers’ and principals’ reservations about NAPLAN, it is clear that the results of these tests are now routinely used by schools in local planning and monitoring of school-level progress.

## Individual teachers

About half of the respondents to the review’s online survey expressed a clear preference about who should have access to NAPLAN results. Among these there was near universal agreement that teachers should have access to detailed individual reports on

students' achievement. Some stakeholders reported high levels of teachers' use of this information. One of the Practitioners' Reference Group members, for example, reported that all teachers in their school had access to the NSW Government's Scout data analytics tool and '90% of staff use it and find it quite user-friendly'. At the other extreme, one of the teachers' union stakeholders reported that 'teachers do not engage with or talk about NAPLAN data'.

The key issue in terms of teachers' use of NAPLAN data concerns the balance between test scores and teachers' judgments. One member of the Practitioners' Reference Group mentioned that "it's nice at times to get an indication of how well our students do in an external test" but others argued that teachers "know where our students are at" and prefer to "use the data they gather from students on a day-to-day basis". Where teachers saw value in NAPLAN data it was typically in combination with teachers' judgements or other assessments – 'It doesn't replace teacher judgement, but it can provide a catalyst to have a conversation with a teacher about a child's progress' (Member of the NAPLAN Review Practitioners' Reference Group), and 'Students are flagged through multiple data sets, including NAPLAN which can help triangulate where a student is in need of support' (Member of the NAPLAN Review Practitioners' Reference Group).

## Family and community

Beyond the professional communities of teachers, schools and school systems/sectors, NAPLAN results are available to parents/carers and the broader community in two ways – through individual students' NAPLAN results and through the publication of school-level NAPLAN results on My School, on school system/sector websites and on individual school websites.

There was widespread appreciation of the individual NAPLAN results provided to families. Parent group stakeholders noted that 'parents/carers appreciate seeing an average that they can compare their child against', that it is 'incredibly important to provide parents/carers with external judgment' about individual student achievement and that NAPLAN 'can complement school reporting'. Principals' association stakeholders reported that 'parents/carers want to see how their child performs in relation to their year group' and that they are 'interested in the data to compare their school's performance and to compare their child's growth and overall achievement with the information the school is providing them'. This was confirmed by members of the Practitioners' Reference Group, one of whom commented that "parents/carers value NAPLAN data" because "it is an external test they can use to compare their kids nationally before the HSC".

In contrast with their use of individual reports, principals reported that few parents/carers mentioned or relied on school-level NAPLAN achievement data. One principal said that in seven years as principal in a low-socioeconomic status (SES) school they "did not have one parent ask about NAPLAN data". Another referred to their experience in several high SES schools where parents/carers were 'quite vocal' but were 'not worried about NAPLAN.' This is consistent with findings of the 2018 Queensland NAPLAN Review Parent Perceptions Report, which concluded that:

Parents were not generally familiar with the content of the tests and tended to be unaware of the full range of NAPLAN reports. Indeed, many parents felt they had not been given clear messages about what NAPLAN is or what it is for (Matters, 2018, pp. 33-34).

The same report noted that a strong majority of parents were 'particularly critical of the role of the media in making NAPLAN a high-stakes assessment through publishing league tables and placing too much emphasis on school results' (Matters, 2018, p. 33).

One of the ways in which NAPLAN results have an impact is on organisations that provide services to schools. Several learning difficulties organisations reported that the release of NAPLAN individual results coincided with an increase in referrals for assessment and inquiries about support. As one of the parent/carer association stakeholders explained, NAPLAN may flag difficulties that have not been made clear to parents/carers by schools.

NAPLAN can also be external evidence of any learning difficulties or if a student is behind. Particularly in primary schools, teachers are reluctant to give out D and E grades. Teachers always used "strength-based language" as well. Parents/carers can be uninformed about their child's performance in relation to the child's cohort without NAPLAN.  
(Parents/carer's association)

Several parent/carer association representatives commented on the role of NAPLAN in school choice. One reported that in their consultation for this review dozens of people responded and 'all said that that NAPLAN results were a consideration when selecting a school'. Others cautioned that there is no value in choice if there is 'only one local school' or if parents/carers 'can no longer school shop' because of school zoning regulations. One of the independent school sector stakeholders drew attention to a previously reported survey that showed only 18% of parents/carers had NAPLAN in their top three reasons for choosing a school. Similarly, a parent/carer stakeholder group argued:

School performance is a secondary consideration. Less than 5 per cent of parents/carers who took part in APC's 2018 Survey, said NAPLAN results were important in choosing a school, for example. Parents/carers make their own judgements about a school based on a range of factors. (Written submission response: Parents/carers' association)

The release of school-level NAPLAN results was much more controversial. More than two-thirds of those who responded to the review's online submission process offered an unambiguous opinion on the public release of data associated with NAPLAN. Of those 67%, 54% were against public release of any data, 6% favoured public release of global data but not school-level data and 7% favoured release of school level data, citing greater transparency as the reason for doing so. The tenor of these concerns about public release of data is captured in this online response to the review:

I believe that data should be provided to schools and parents/carers but not the media. We have enough to deal with in schools without having to worry about comparisons which the public make between schools based on NAPLAN data.

Several sources report, however, that despite this low level of use most parents/carers agreed that school-level NAPLAN results should be available on a public website (Louden, 2019, p. 91). Similar conclusions about parents/carers' perceptions have been reported in a recent review commissioned by ACARA.

Parents had generally not given this issue prior thought and, when questioned, found it difficult to see a reason why the information on the My School website would not be freely available in the public domain (ACARA, 2018b, p. 5).

## Summary

For more than a decade, NAPLAN's standardised assessment data have underpinned public reporting on national, state and territory trends in student achievement and growth. Wide differences in school-level achievement and growth may be observed among schools serving similar communities, and these differences are reported to schools through systems/sectors' data analytics tools and to the public through My School and school system/sector websites.

This review's evidence of the uses to which NAPLAN results and analyses are put is broadly consistent with other recent reviews. Stakeholders confirmed that states and territories set targets for NAPLAN achievement and that NAPLAN results are routinely used by school systems/sectors and schools in target-setting, planning and monitoring achievement. Many stakeholders, however, were concerned that school-level NAPLAN targets may narrow teachers' focus to students near the boundaries of bands of achievement and unfairly categorise schools with lower than expected student achievement. Similarly, the 2018 Queensland NAPLAN Review concluded that NAPLAN had been 'both a positive and negative driver of education' (Cumming et al, p. 14). Participants 'were comfortable with educational accountability for transparency of educational outcomes and monitoring the health of an education system' but concerned that 'emphasis on NAPLAN as an accountability measure at system and school levels continues to create a negative competitive environment for systems and schools, perpetuating negative educational practices in some schools' (p.14).

Teachers and principals responding to this review indicated that school-level NAPLAN results were used to track trends in cohorts and identify skills that may need more focus, but that results at the individual level were of less use. Where teachers saw value in NAPLAN data it was typically in combination with teachers' judgements or other assessments. This is consistent with the Queensland NAPLAN Review which reported 'extensive data collection in schools for triangulation, including NAPLAN data' but 'limited engagement with NAPLAN test data' (Cumming et al, pp. 11-12).

Principals reported that parents/carers rarely mention or use school-level NAPLAN results, a conclusion confirmed by the Queensland NAPLAN Review Parent Perceptions Report (Matters, 2018). Parents/carers, principals and teachers consulted in this review, however, reported that families appreciated the individual-level NAPLAN results because they provide a judgment external to the local school.

The most controversial use of NAPLAN results is the publication of school-level NAPLAN results. Many professional stakeholders are opposed to the publication of school-level NAPLAN results because they can be used for school comparisons and league tables. There is, however, a tension between this view and evidence of broader community expectations about transparency of school-level achievement data. Few parents/carers consulted in either of the recent NAPLAN reporting reviews (ACARA, 2018; Loudon, 2019) had used the school level reports NAPLAN reports available on the My School website but nevertheless most agreed that school-level NAPLAN results should be available to the public.



# Chapter 7: Recommendations

This final chapter sets out the rationale for specific changes proposed in the recommendations offered. It relates the recommendations to the review terms of reference and proposes a timeline for implementation.

## National standardised assessment

Standardised assessment provides one way to see how well education is progressing. The population may ask this of the whole system/sector. Parents/carers may ask it of their own children or their children's school. Using common test-taking conditions, questions, time to respond and scoring procedures, standardised assessments can provide answers framed in a larger perspective than local classroom or school assessments can provide.

### Purposes of national standardised assessment

Chapter 1 identified the following five important purposes for national standardised assessment that have been endorsed through a decade of decisions by national ministerial councils in Australia.

### Monitoring progress towards national goals

Measurement of students' achievements and progress over time can monitor the progress of the education system toward national goals at national, jurisdictional and system levels. They can also provide information on the relative performance of students by gender, geographic location of schools, socioeconomic background and Indigenous background. This can be achieved with domestic data collections or through Australia's participation in international surveys such as the Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA).

### School system accountability and performance

Measurement of students' achievements to monitor progress towards national goals can also provide public information on school system accountability, including inter-jurisdictional and inter-sectoral comparisons and information on the performance of students in equity groups.

### School improvement

The Australian results from the international PIRLS, TIMSS and PISA surveys can identify areas of general weakness in Australia as a whole or in particular states and territories to which schools can respond but, except for schools in the sample, it would not be their own students' data that they would be examining. For detailed comparative data to which all schools can respond, all students need to be tested in a census.

### Individual student learning achievement and growth

A focus on individual student achievement and growth requires individual students to be tested. The data can be obtained simultaneously by testing all students at the same stage of school or by testing small groups or individuals at the discretion of schools or individual teachers. Simultaneous testing allows comparison and interpretation of individual student's performances in the light of the performances of the whole population or relevant sub-populations. When individuals or groups of students are tested with published standardised tests, there will usually be norms for the relevant age group or year level to provide comparative information. The comparisons

would not be precise if the norms were determined at a different time in the school year from the time that the school uses the test.

### Information for parents/carers on school and student performance

Standardised tests can provide information for parents/carers that is situated in a wider frame of reference than their children's own school. It is this wider frame of reference that enables parents/carers to have some understanding of the position of their own children in relation to the population of which their children are part and, depending on the scope of the data and the extent of their access to it, some understanding

of the school's position among all schools or among schools with students similar to their own. The frame of reference could be provided by results for the relevant population of students obtained at the same time on the same standardised tests or from norms established on the test at an earlier time with similar students.

## Features of an assessment system

### Sample versus census testing

Census tests have the capacity to meet a wider range of the purposes of a national standardised assessment program than do sample tests, as shown in the summary in Table 23.

**Table 23:** Census and sample assessment and the purposes of national standardised assessment

Purpose of national standardised assessment	Census	Sample
Monitoring progress towards national goals		
• National, jurisdictional and system estimates of achievement	✓	✓
• Relative performance by gender, geographic location of schools, socioeconomic background and Aboriginal and Torres Strait Islander background	✓	✓
School system accountability and performance		
• Accountability for system performance	✓	✓
• Accountability for school performance	✓	✗
School improvement		
• School-level information on achievement and growth by assessment domain	✓	✗
• School-level targets informed by system comparative data	✓	✗
Individual student learning achievement and growth		
• Student level achievement estimates for comparative purposes (cohort, test domain, gain, equity groups)	✓	✗
• Student level achievement estimates for diagnostic purposes	✗	✗
Information for parents/carers on school and student performance		
• Individual student achievement	✓	✗
• Relative school performance	✓	✗

Census tests can be an appropriate source of information for monitoring national policy, system accountability, school improvement and reporting to parents/carers on school and individual performance. Although they have less diagnostic value at the individual level than more intensive and extensive standardised assessments designed for diagnostic purposes, they can provide a 'point-in-time' indication of a student's position in relation to the whole population of which the student is a member. Students' results can be reported directly to parents/carers. They may also signal the need for exploration with specialised diagnostic assessment.

Sample tests, on the other hand, are effective only in monitoring progress towards national goals and accountability for system performance. Sample tests have the benefit of reducing the risk of some of the unintended consequences of census testing. They have a lesser tendency to narrow the curriculum in schools because only some schools are involved and because a wider range of knowledge, understanding and skills can be measured when only a sample of students is involved. They do not enable school-by-school statistical comparisons that many, particularly teachers, find undesirable but they also reduce transparency, limit school-level accountability and invite the imposition of other census test regimes to support school systems/sectors' school improvement targets.

Stakeholders' views on NAPLAN were sought through two rounds of interviews, a written submission process and the opportunity to complete an on-line survey. To obtain more in-depth practitioner perspectives, meetings were also held with a Practitioners' Reference Group, including principals and teachers from government, Catholic and independent schools across the four participating jurisdictions, and a nominee of the Australian Education Union. More

than 300 people chose to respond through the online submission process and about half of these people had a clear view on whether national standardised assessment should be based on a sample or use whole-population census testing. Of these, about half (25% of the total respondents) supported continuation of census testing. Their reasons included 'the opportunity for all parents/carers to gain information about the progress of their child', the ability of schools to 'determine strengths or weaknesses of individuals, cohorts, and the whole school over time', and concern that sample testing would mean that 'trends could be dismissed as sampling errors'.

A similar proportion (22% of the total respondents) advocated for a move to sample testing. These respondents argued that sample testing would reduce pressures to teach to the test and eliminate school-by-school comparisons but maintain the value of national assessments as 'a health check for the education system as a whole'. As one respondent put it:

While some data would be lost by moving to a sample system, it would certainly take the pressure off schools. System wide data might refocus efforts on equity measures, rather than punitive targeting of individual schools. (Respondent to the online survey)

Among the half of respondents who did not support either sample or census testing, the largest proportion (19% of total respondents) were opposed to NAPLAN in principle, opposed to high stakes testing more generally, or had concerns about the reliability of the current tests. Of the remainder, some respondents expressed concern about sampling error (12%) and others were ambivalent, expressing views both for and against sample testing (9%).

Stakeholder interviews revealed the same broad range of views. School system/sector stakeholders most often supported continuation of census testing. One of the large non-government school systems/sectors, for example, acknowledged that sample testing would reduce unintended consequences such as test anxiety and teaching to the test, and would be well-received by some schools, but noted that:

one of the disadvantages of sample testing is that it removes the main benefit that NAPLAN data currently provides to schools, being that data is provided for all students in Years 3, 5, 7 and 9 enabling subsequent direct comparisons and the identification of learning growth trends over time. (School system/sector)

School system/sector representatives noted the greater analytic power of census tests, warned against the loss of systems/sectors' evidence base, the loss of their schools' capacity to examine progress over time and the loss of universally comparable individual student reports to parents/carers. Moreover, as one system/sector representative warned, in the absence of whole-population assessment "the void may be filled with something else" because school systems/sectors' "big picture school improvement work wouldn't be possible with a sample test".

Among teacher union stakeholders, there was universal preference for sample over census testing. Union stakeholders acknowledged the legitimate role of national standardised testing in monitoring the performance of school systems and the targeting of resources to equity groups, however they argued this could be achieved without what they characterised as the negative consequences of census testing – student stress, narrowing the curriculum, teaching to the test and comparisons of schools on the My School website. As one of the written responses argued:

The legitimate needs of system self-monitoring can be met by representative sampling methods which can provide accurate and useful information without any of the negative outcomes of mass standardised testing. This would give an overall snapshot of student achievement in each state and territory jurisdiction... and enable education authorities to track the progress of various student cohorts such as Aboriginal and Torres Strait Islander students... (Written submission response: Union)

The views of other stakeholders were more mixed. Principals' associations typically preferred sample testing. Some subject association stakeholders preferred sample tests, others preferred census tests and one stakeholder was 'more sample-orientated but could be convinced of census tests'. Among the members of the NAPLAN Review's Practitioners' Reference Group there was some support for the school-level data available from NAPLAN census tests but more often the sentiment was to support a move towards PISA-style sample tests. Some parents' association representatives preferred census testing, noting that 'census testing works well', that 'schools are doing things about their kids learning because of the data and it is being useful' and that, without census testing parents/carers, 'would not get an external assessment on their child'. Another parent group stakeholder preferred sample testing but cautioned that "not having information on every school wouldn't 'wash' with Ministers".

Among educational experts responding to the review, judgements about whether to prefer sample to census testing typically turned on the question of purposes of assessment. They noted that census testing 'allows for accountability and reporting to parents/carers' and 'exposes disadvantage', but that census data are 'less precise

for individuals' than groups. Sample testing using longer test instruments would increase precision and 'provide the opportunity to test students about more of the curriculum'. Several experts commented, however, that the current National Assessment Program sample testing in scientific literacy, civics and citizenship and information and communication technology literacy does not have an impact in schools. As one said, the reports "make headlines but states and systems do not take action on the basis of the data".

### **Practices in other countries**

The international comparisons provided in Chapter 3, address the issue of sample versus census testing in the various national contexts. In Singapore, the Primary School Leaving Examination (PSLE) provides census testing at the end of primary school with oral and listening and comprehension examinations in English Language and Mother Tongue, and written examinations of one to two hours in English, Mother Tongue, mathematics and science. In middle secondary school, the Singapore-Cambridge General Certificate of Education provides census testing in examinations, depending on which course of study the students are pursuing. In Japan, students take competitive subject examinations at the end of Year 9 for selection entry to senior high schools. In Ontario, there is census testing of students in reading, writing and mathematics at the end of Grades 3 and 6. In England, schools are obliged to report teacher judgements in reading, writing, mathematics and science. A formerly optional English grammar and punctuation test at the end of Key Stage 1 (Year 2) will remain optional but a compulsory, online 'multiplication tables check' is scheduled for introduction in 2019-20. At the end of Key Stage 2 (Year 6) and Key Stage 3 (Year 9) there are census tests in English, mathematics and science. At the end of

Key Stage 4 (Year 11), there are national subject-based examinations for the General Certificate of Secondary Education (GCSE).

New Zealand only has national surveys of student learning but provides access for schools to a range of standardised assessments to assess their own students. In Finland, prior to national examinations at the end of secondary education, assessments of students' progress are school-based but there are sample surveys with standardised tests of students' achievements in particular subjects selected on a cycle. In Scotland, there are census assessments in Years 1, 4, 7 and 10 in reading/literacy, writing and numeracy. While schools have the right to opt out, student participation rates match those achieved by the NAPLAN census assessment in Australia. Students' results are reported only to schools where they are to be used in conjunction with teachers' assessments to create reports on students that are provided to parent, students and local education authorities.

### **Uncertainty in measurement**

There is always a level of uncertainty or imprecision in measurement. Some of it is due to the test itself, some due to the uniqueness of the random sample chosen if only a sample of students is tested (that is, a sampling effect), and some due to links to previous tests if trends over time are measured and reported. As shown in Chapter 4, the level of uncertainty depends on the amount of data behind the measure. There is greatest precision with national means and least with individual student's results. There is greater precision with means for large schools than with means for smaller schools. There are ways in which the degree of uncertainty can be reported numerically or graphically to reduce the risk of over-interpretation of a single individual score or group mean.

## Role of NAPLAN in meeting national purposes

Chapter 6 provides information on the current uses of NAPLAN by governments, education systems/sectors, schools, teachers and parents/carers. It is a somewhat mixed picture reflecting the diversity of stakeholder views about NAPLAN. As reported in the preceding section, there is general support for all the purposes of national standardised assessment shown in (page 114) that can be supported by sample testing but less for those that require census testing. The strongest rejection of census testing was driven by concern about the public exposure of school results and the facilitation of inter-school comparisons that this enables. The culprits included media that produced league tables of schools, taking no account of difference in school contexts, but the main culprit was said to be My School.

The My School website originally provided comparisons only among schools that enrolled students with similar levels of socio-educational advantage but the availability of the results for all schools enabled users to make other comparisons. Among respondents, there seemed to be little awareness yet of the significant changes to My School in 2019 (Louden, 2019) which, as pointed out in Chapter 6, removed the comparison of the achievements in schools with similar students and introduced a

comparison of the improvement achieved between successive NAPLAN tests (Years 3 to 5 and Years 7 to 9) by students in a selected school with the improvement by other students across the country who had the same NAPLAN score two years earlier and who have a similar background.

The concerns about publicly available school-level data are not only that they permit inter-school comparisons but also that the comparisons are limited to the particular student achievements in literacy and numeracy that NAPLAN measures.

There are also concerns that government education systems set state-wide targets in terms of NAPLAN achievement and improvement and set school-level targets based on them. Concern about the impact on schools and teachers of such target setting is behind the kinds of comments this review has heard about “pressure” and “punitive targets”. These are matters about managing improvement rather than about the measures that provide the criteria for setting and judging the improvement. As one jurisdiction’s business analytics tool puts it, NAPLAN census data provide a ready source of information about “what’s working well, and what can be improved”. Jurisdictions have made substantial investments in such business analytics tools for use by the central and regional administrations and schools.

While the absence of school-level reporting would eliminate the possibility of school comparisons, it would not meet the accountability commitment made by education ministers in their 2019 Alice Springs Declaration:

For schools, Australian Governments provide assessment results that are publicly available at the school,

sector and jurisdiction level to ensure accountability and provide sufficient information to parents, carers, families, the broader community, researchers, policy makers and governments to make informed decisions based on evidence. (Education Council, 2019a, p. 11).

## Recommendation 1

- 1.1 Ministers re-endorse the importance of standardised testing in Australian school education for:
  - a. Monitoring progress towards national goals.
  - b. School system accountability and performance.
  - c. School improvement.
  - d. Individual student learning achievement and growth, noting the limitations on use in detailed diagnosis of learning deficiencies and difficulties due to the degree of uncertainty in measures of individual students.
  - e. Information for parents/carers on student and school performance.
- 1.2 Ministers re-affirm the role of national standardised assessment in fulfilling these purposes.
- 1.3 Continue to conduct national standardised assessment as a census test of student achievement.
- 1.4 Define the purposes and limitations of national standardised assessment by decision of the Ministerial Council in the manner proposed in Table 23 and communicate this on the Australian Curriculum, Assessment and Reporting Authority (ACARA) website and in communications with schools and parents/carers.

## Changes to the NAPLAN tests

### Curriculum coverage

#### Connection with the Australian Curriculum

As discussed in Chapter 1, Australian students used to take subject-based assessments as external examinations at the end of primary school and in mid-secondary school. Now they take no such assessments before the end of secondary school.

Over a period from the late 1980s, states and territories introduced new external assessments to be taken by all students at various stages prior to the end of secondary school before, in 2007, the ministerial council resolved to replace the external assessments from 2008 with the common national assessments known as NAPLAN.

All these assessments, both NAPLAN and the state and territory assessments that preceded it, have been limited to literacy and numeracy. There was no desire to reintroduce extensive subject-based assessments for all students, so the focus was placed on the foundational domains of literacy and numeracy on which so much other learning depends. There are concerns, however, that the focus on literacy and numeracy has had unintended effects.

An important concern is that it has narrowed the curriculum through too much attention being given to literacy and numeracy at the expense of other things that ought to be central to students' learning and their general experience of school. This narrowing effect is seen to have had a greater impact on primary schools since other subjects are protected in secondary school through specialist teachers and timetables that allocate space for all subjects.

An extensive focus on literacy and numeracy in primary schools is not new. In a study for the Australian Primary Principals Association of time allocations in the primary school curriculum, Angus, Olney & Ainley (2007) reported that, 'Up to the twentieth century,

the elementary school curriculum was truly elementary: over three quarters of the time was spent on literacy and numeracy' (p.15). Despite a broadening of the curriculum and complaints that it had become overcrowded, their survey of teachers' actual time allocations revealed that 'One of the realities of primary schools is that more than half the instructional time is spent on English and Mathematics' (p.24).

As Chapter 4 reported, many stakeholders expressed concern about a lack of alignment between NAPLAN and the Australian Curriculum. There has, however, been a substantial effort to ensure that this is not the case. ACARA acknowledges that 'NAPLAN draws on all learning areas of the Australian Curriculum to supply contexts for testing' but 'they do not assess the content of learning areas other than English and Mathematics' (ACARA, 2017, p. 6). The reading test is designed to 'assess students' ability to read and view texts to identify, analyse and evaluate information and ideas' and is aligned to English.

As set out in the Australian Curriculum: English, students read texts for different purposes: personal interest and pleasure, to participate in society, and to learn. Since the emergence of visual and digital communication media, the traditional view of literacy has broadened and evolved, and viewing is now a key literacy skill. NAPLAN assesses students' ability to read and view multimodal texts for literacy experience and to acquire, use and evaluate information. (ACARA, 2017, p. 9)

Similarly, the spelling tests draw on the spelling sub-strand; grammar draws on the sub-strand threads of text cohesion, sentences, and clause and word level grammar; and writing draws on seven sub-strand threads of the Australian Curriculum: English (2017, p. 15). The curriculum connection is even closer between numeracy and the Australian Curriculum: Mathematics, as the tests draw on both



the proficiency strands (understanding, fluency, problem solving and reasoning) and the content strands (number and algebra, measurements and geometry, and probability and statistics) of the Australian Curriculum: Mathematics. The connections should have been clear through the naming of reading, language conventions and writing NAPLAN tests, but for the avoidance of doubt there is an opportunity to make the link clearer by renaming numeracy as mathematics. There is no case for renaming the rest as English since reading and writing and language conventions are named and tested separately.

### **Concern about teaching to the tests**

Another aspect of the concern about narrowing of the curriculum is that it will become limited to those aspects of literacy and numeracy that are actually tested in NAPLAN and that teachers will allocate time unproductively to teaching to the test and to unnecessary student practice in taking NAPLAN-like tests. This effect has been most obvious in writing in which students seem to learn formulaic ways of writing in particular genres in response to the prompts in the NAPLAN writing test but it is said to occur in all NAPLAN test domains.

The branching structure of the NAPLAN Online tests provides some protection against teaching to the test because there is no single test in each domain that all students take. Even those students who follow the same path through the branching structure will not necessarily be responding to the same questions. As noted in Chapter 4, in the 2019 NAPLAN Online reading and numeracy tests, for which there were seven pathways through the testlets, there were actually 126 paths through the actual test items. If the items effectively cover the Australian Curriculum: English, the Australian Curriculum: Mathematics and the literacy and numeracy continua in the Australian Curriculum, the only effective way to prepare students for the

tests would be through implementing the Australian Curriculum.

A further influence on narrowing the curriculum is reported to be the publication of school results in NAPLAN tests, particularly on My School, but also as a requirement in school reports. Publication is also said to narrow the conception of quality of schooling and invite judgements of school quality on the primary basis of literacy and numeracy results.

Results of the National Assessment Program surveys of samples of Years 6 and 10 students in science literacy, civics and citizenship, and information and communication technology (ICT) literacy are published but do not seem to have much influence on public discussion of the performance of the education systems/sectors. This is despite the reports providing comparisons among states and territories and analyses of the relative achievement levels of sub-populations of interest.

### **Inclusiveness of the tests**

The NAPLAN tests cater well for students whose special needs require adjustments to the form and delivery of the tests. The concern is not with the inclusiveness of the tests but the extent of coverage of the full student cohort in two respects. First, some parents of students with learning difficulties were disappointed that their children's schools had urged that their children be excluded from the tests. These parents had not exercised the option to withdraw their children. Rather, the school had placed them explicitly in the exempt category or, by default, in the absent category. Secondly, there are numbers of students who are simply absent on the days of testing to an extent that exceeds those missing through exemptions or approved withdrawal. The absence rates should be investigated to learn why students do not arrive for the assessments with action then taken to reduce the rates.

## Broadening the range of the tests

It is timely to consider broadening the range of NAPLAN tests. Literacy and numeracy have been the focus because they are foundational and in order to restrict the scope and impact of census testing. The international assessments in PISA and TIMSS include science in which Australian students' relative and absolute achievement levels have been declining. Japan and England include science with national language and mathematics in census testing. Singapore includes science as one of the subjects in its primary school leaving examination at the end of P6 (Year 6). In Australia, all jurisdictions are placing new emphasis, not only on science, but on STEM more generally. In terms of the Australian Curriculum that would incorporate not only mathematics under the 'M' but digital technologies under the 'T' and 'E'.

There is also increased interest in Australia in learning and assessment of the General

Capabilities in the Australian Curriculum. One of them at least, critical and creative thinking, is fairly clearly domain specific, particularly the critical thinking aspect. Critical and creative thinking in history is not the same as critical and creative thinking in STEM, for example, so the assessment would need to be situated in a domain. Adding critical and creative thinking in STEM to Australia's census assessments would reflect the priority being attached to both STEM and the General Capabilities.

It would take time and a deal of experimentation to develop the new assessments and it would be best not to attempt to apply them at Year 3, at least in the first instance, until the nature and scope of the assessments are well-defined and valid and reliable tests have been developed. It could be replaced in the triennial cycle of sample surveys with a new assessment such as in history and intercultural understanding.

## Recommendation 2

- 2.1 Ministers note that national assessment policies and practices vary and that there are no common features of assessment among high-achieving countries.
- 2.2 Rename the numeracy test as mathematics, to clarify that it assesses the content and proficiency strands of the Australian Curriculum: Mathematics.
- 2.3 Add assessment of critical and creative thinking in science, technology, engineering and mathematics (STEM) to the national standardised census assessment program, except at Year 3, and introduce it only after a period of experimental test development that demonstrates that valid and reliable tests have been developed.
- 2.4 Withdraw the current triennial sample survey of science literacy in Years 6 and 10 and consider replacing it in the triennial cycle with another covering both a subject and a general capability from the Australian Curriculum, such as history and intercultural understanding.
- 2.5 Explicitly map the tests to the National Literacy and Numeracy Learning Progressions to provide insight into student learning progress in the year levels in which the tests are administered.
- 2.6 Jurisdictions investigate students' reasons for absence from NAPLAN testing and seek to reduce the current levels of absence, particularly at the secondary level.

## Frequency and timing of tests

From 2008, when the states and territories adopted NAPLAN as common, national tests, NAPLAN has been administered in May to all students in Years 3, 5, 7 and 9.

### Timing of testing and return of results within the year

The NAPLAN Review interim report raised the possibility of shifting the timing of the testing from May to late February or early March, based on the following considerations.

Shifting the tests to early in the year, combined with speedy delivery of results, would make NAPLAN a measure of teachers' and students' starting points for the year. It could liberate NAPLAN to play a formative rather than a summative assessment role and to inform decisions about future curriculum and teaching choices, not judgements about past ones.

It is, of course, possible that start-of-year assessments would be seen as summative assessments of the end of the previous year. That argument is potentially weakened by the impact of declines in student performance over the summer vacation and the tendency for class groups to be formed with different mixes of students in each new year of schooling.

Assessment of the starting points for the year could give school systems the opportunity to provide additional resources to schools in most need of additional support (McGaw, Loudon & Wyatt-Smith, 2019, p.6).

In submissions to the review and in consultations, there was general support for this proposal. The only considerations in determining how early in the school year the tests could be administered, would be how long it would take for students to be settled into their new classes and for schools to settle the class rolls.

Administering NAPLAN early in the school year would reduce the likelihood that any school might spend time preparing students for the tests beyond ensuring familiarity with the format. Once NAPLAN Online is fully implemented, results could be returned to schools and students within days of testing. That would reinforce their value as a measure of the starting point of the year.

### Years of testing

This review's interim report also raised the possibility of the tests being taken by students in other Years than 3, 5, 7 and 9. The primary considerations were whether to shift from 3 and 7 and, if so, to adjust the other years or to eliminate them. Testing every two years would provide a better view of students' growth over the school years than would testing on only two occasions. On the question of when to start, the review's interim report canvassed the possibilities of Years 2, 3 or 4 for the first tests.

Would Year 3 be too early if the tests were at the beginning of the school year? On the other hand, would waiting until the beginning of Year 4 be too late, given the importance for students' academic self-concept of becoming secure readers early in their school lives?

In submissions to the review and in consultations, both of these options and also the possibility of testing in Year 2 were raised. Year 2 would have the benefit of earlier detection of emerging learning problems for students, but NAPLAN-style tests would not be appropriate for students at that age. Furthermore, with all systems/sectors now using early screening by one means or another, at least in literacy, early detection should already be in hand.

There was also some support for testing in Year 6 rather than Year 7. That would make the assessment more summative for primary schools, or for the primary stage in Foundation to Year 12 schools, and that

would reinforce the judgemental role of NAPLAN that worries those who think it provides too narrow a set of criteria for evaluating schools. On the other hand, keeping the testing in Year 7 and moving it to early in the school year, would give secondary schools formative information about their incoming students. Many secondary schools have students take standardised tests early in Year 7, or even in Year 6 if they know which students will be coming to them in the new year. Respondents from schools with these practices claimed that they did not believe that reports on students from the various primary schools provided comparable assessments. If NAPLAN were to test all students early in Year 7, secondary schools may well be able to abandon other assessments that they currently use at the beginning of the year or late in the prior year.

The possibility of shifting from Years 7 and 9 to Year 8 and 10 was raised in the review's interim report for consideration (McGaw, Loudon & Wyatt-Smith, 2019, p.4). An alternative would be to maintain testing in Year 7, at the commencement of the secondary school years, but to delay the later testing to Year 10. Year 9 is generally regarded as a difficult year for students and schools and the NAPLAN test results certainly reveal a low level of engagement of Year 9 students. With Year 10 a key year for students' decisions about their future study options, the commencement of that Year would be a good time to obtain for the students, their teachers and their parents/carers an assessment of their current progress.

### **Recommendation 3**

- 3.1 Conduct NAPLAN tests as early in the school year as is administratively feasible.
- 3.2 Set a goal for the results from all NAPLAN Online tests marked online being reported to schools, students and parents/carers within a week of the conclusion of the testing window.
- 3.3 Continue to administer NAPLAN tests in Years 3, 5 and 7 and replace assessments in Year 9 with assessments in Year 10. Assessments in Year 9 are not to be held in 2021.

## Rebranding the program

The current NAPLAN census tests are part of a broader National Assessment Program that includes sample surveys in science, civics and citizenship and information and communication technology literacy. This review has proposed a rebalancing of the sample and census assessments, with reading and numeracy to continue as census tests, critical and creative thinking in science, technology, engineering and mathematics (STEM) to be added as a census test from Year 5; and writing to be rebuilt -- first developed as a sample test and later implemented as a census test.

The addition of a census test in critical and creative thinking in STEM would lead to an overlap with the current sample survey in science, which should, therefore, be discontinued. Instead, a sample survey in some other domain, perhaps history in combination with a general capability such as intercultural understanding, could be added to the three-year sample survey cycle.

At present the National Assessment Program (NAP) is an umbrella title for both the sample surveys and the literacy and numeracy census tests as NAPLAN. To distinguish them more clearly and to recognise that the census tests are proposed to move beyond literacy and numeracy, it is proposed that new names be adopted for each program: the Australian National Standardised Assessments (ANSA) instead of NAPLAN and National Sample Assessment Program (NSAP) instead of NAP.

### Recommendation 4

- 4.1 Adopt a new name, Australian National Standardised Assessments (ANSA), in recognition of the changes in the existing tests and the addition of tests of critical and creative thinking in science, technology, engineering and mathematics (STEM).
- 4.2 Discontinue the National Assessment Program (NAP) sample survey in science literacy with the introduction of ANSA in critical and creative thinking in STEM.
- 4.3 Maintain the National Assessment Program (NAP) sample surveys in civics and citizenship and in information and communication technology literacy on their current three-yearly cycle. Rename the program the National Sample Assessment Program (NSAP).

## Redeveloping the online branching tests

### Creating new 'digital-native' tests

Completing the move to online tests in the absence of simultaneous use of print versions of the tests will liberate the online tests from the restriction imposed by a requirement to parallel the print versions. The new online tests should be 'born digital' so there should be no requirement to match the online versions used in the transition period when parallel print and digital forms are used.

The new digital tests will require creative test development to capitalise on the flexibility and capacity of digital delivery. The stems for test items need not be static and responses could be constructive. Simulations could be used, for example, to engage the students' in complex analysis and reflection.

There are two benefits of a move to exclusive use of digital tests for vertical equating of the scales over Years 3, 5, 7 and 10 and horizontal equating of the scales over years of testing. First, the complications faced in equating both print and digital versions in the transition years will be avoided. Secondly, provided a new time series is started and links back to 2008 are abandoned, there will be no need to use common-person equating with the secure print tests from 2008 in the horizontal equating. All the equating will become common-item equating. The earlier decision to cease publishing the NAPLAN items will make more items available for repeated use over years and that will enable the vertical and horizontal links between tests to be strengthened.

There will be lessons to be learned about equating in this new arrangement. As described in Chapter 4, each testlet in the branching tests for Literacy and Numeracy has three forms that need to be parallel in the sense of having the same breadth and depth of coverage of the curriculum and

equivalent item difficulty levels. If this is not achieved, whether students on the same path AD, for example, might be directed then to F or E (see Figure 2, p. 60), could become an arbitrary consequence of the particular versions of the testlets A and D with which they were presented. Technically, this should not matter because of the claim that the psychometric model used can establish the students' achievement levels on the NAPLAN scales independent of the difficulties of the particular items to which they responded. The difficulty is that the process at present uses number of items correct and not calibrated scores on the testlets to determine which branch to take, so it is essential that the three versions of each testlet have very similar item difficulties. To minimise these risks, it will be important for there to be thorough trialling of the test items before they are used in NAPLAN Online tests. There will need to be trialling to confirm that the branching based on determined item difficulties works effectively as well as prior trialling to establish the item difficulties. There should also be further development of the computer delivery platform to see if the branching could be based on calibrated estimates of achievement derived by the psychometric model and not number of items answered correctly in the testlets.

### Setting new benchmarks

The discussion of benchmarks in Chapter 4 (see Table 16, p. 75), where it is clear that fewer students fail to reach the NAPLAN benchmarks than fail to reach the other benchmarks. This could be because the students are better prepared for the NAPLAN tests because of their connection with the Australian Curriculum or engage with them more because they are domestic census tests rather than international sample surveys. The other obvious possibility is that the NAPLAN National Minimum Standards (NMS) benchmarks are less demanding than the international ones.

That is even more likely given that all students who did not sit the NAPLAN tests because they were exempt are counted as below NMS. The percentage of those who sat and are below NMS is therefore smaller than the percentages shown in Table 16, p. 75.

Table 16, p. 75, revealed that a smaller proportion of Australian students fall below the minimum standard thresholds in the NAPLAN tests than fall below the minimum standard thresholds set in the international surveys of student achievement in which Australia participates, Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA). Differences among the various surveys in

what is tested are discussed in Chapter 2 but, regardless of those differences and the possibility that lower proportion below the minimum standard in NAPLAN is due to NAPLAN being more closely aligned with the Australian Curriculum, the levels at which minimum competence in NAPLAN are set should be reviewed. Work should also proceed on the development of 'proficient' and 'highly proficient' NAPLAN benchmarks.

## Recommendation 5

- 5.1 Redevelop the reading and mathematics tests as digital assessments, capitalising on all the flexibility that the digital form offers for content and item form, with no constraint to mirror the current print versions of the tests.
- 5.2 Develop the new critical and creative thinking in science, technology, engineering and mathematics (STEM) so that it is 'born digital' since it will have no print form that it might have been constrained to match.
- 5.3 Undertake further development of the branching model and system changes to see if the branching could be based on estimates of achievement derived by the psychometric model and not number of items answered correctly in the testlets.
- 5.4 Review the level of the National Minimum Standards on the NAPLAN scales to see if they are set too low and progress work on developing additional 'proficient' and 'highly proficient' benchmarks.

## Redeveloping the writing test

As indicated in Chapter 5, the writing test attracted the most sustained negative comment from stakeholders in this review. Throughout the consultations there were sustained calls for change to the current approach to testing writing in NAPLAN. These calls go well beyond adding another form (for example, informative writing). Among the most common issues raised were – the need for using richer prompts, broadening the range of forms or genres; examining the criteria and accompanying scores against which writing is assessed; changing the conditions in which students are required to write to permit time for planning and review in composing processes; and the potential benefits of including a component of teacher judgement, beyond involving teachers in state-based NAPLAN Marker Quality Teams.

It is interesting to note that while cohort gain is a feature of the National Report analyses of NAPLAN reading and numeracy data, cohort gain for writing is not included. Further, the panel was advised that the writing results are used to a lesser extent by systems/sectors and schools than results for reading and numeracy. An explanation for this lack of use of the writing test results included the following observations:

- Writing results are perceived to be less reliable than those for reading and numeracy, as they are subject to more external sources of variance than other test data.
- These sources consist of genre effects, prompt specific effects, marking criteria and marking consistency.
- Collectively, these external sources of variance can have a significant undue influence on the trends of writing results. For example, Figure 11 shows that, with few exceptions, results of all jurisdictions move in unison from one year to the next, likely to reflect the influence of common external sources of variance (for example,

prompt and equating effects), rather than any real changes in the states and territories underlying performance over time. (Measurement expert)

These deficiencies could be dealt with if the test was fully redeveloped to offer a broadened range of prompts and forms of writing, with altered test conditions, and explicit provision for student choice in how the test is designed.

### **Purpose of the NAPLAN Writing assessment is not clear**

Currently the relationships of the writing assessment and the marking criteria to the Australian Curriculum: English and Achievement Standards, General Capabilities, and the National Literacy and Numeracy Learning Progressions remain opaque.

Clarity about the relationships is essential so that teachers do not see NAPLAN writing in isolation from their classroom practice. It would support teachers' diagnosis of students' learning needs, as well as their selection of curriculum adjustments to support students with disability. With the move to NAPLAN Online, exemplars of student writing, with rich annotations, could be provided to illustrate student writing at different year levels and in NAPLAN performance bands. This would all help build teachers' assessment literacy.

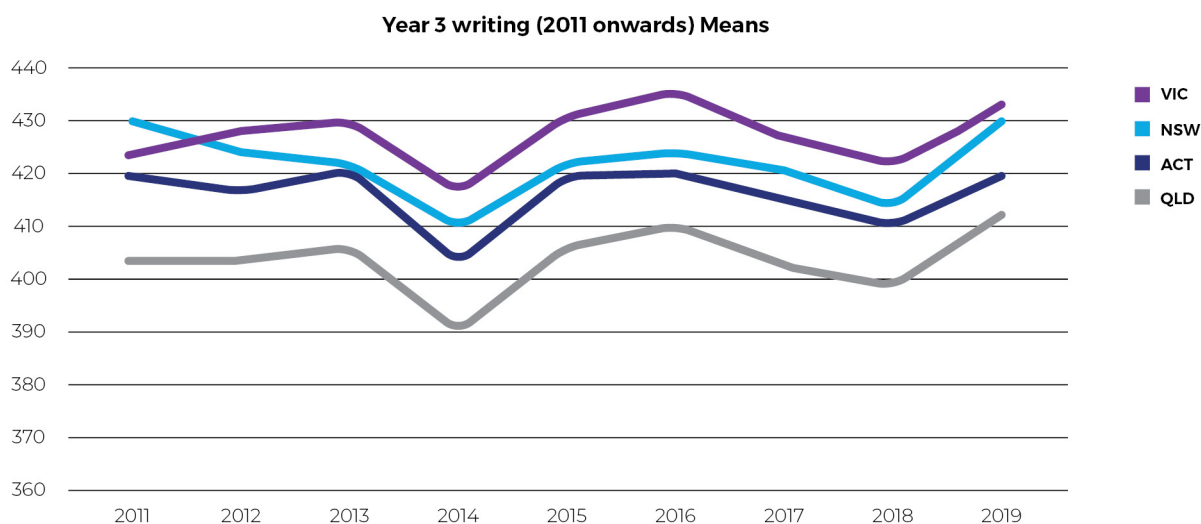
### **Tasks are often decontextualised without sense of audience and purpose for writing**

Teachers frequently mentioned that emphasis on audience and purpose is integral to how students are taught writing in the classroom but largely absent from the NAPLAN writing test, which was characterised as 'alien'. The Australian Curriculum: English requires students to manipulate language features appropriate to audience and purpose but where 'the audience is generic, students do not have a sense of the level of formality required' (Educational organisation).



When students write in school, they have time for planning, drafting and editing and the prompts for their writing generally cue them effectively to select vocabulary and

language features to build an effective relationship with the reader. The NAPLAN test does not enable most to produce their best writing.



**Figure 11:** Trends in mean performances on the NAPLAN Writing test

### Writing required is too narrow

A new writing test could include a wider range of prompts designed to take account of developmental stages of writing. The scope of the writing required could also be broadened to include an extended response or multiple short writing tasks or both, subject to testing purpose and scheduling. Short writing tasks could produce useful information about students' writing skills, particularly in secondary education where the informative genres are relevant. The length of the extended response could be reviewed for Years 7 and 10 and considered in relation to time allocated to test implementation. The extended response and multiple short writing pieces could be staged, if two test window opportunities were available.

At present a single genre for responses is used each year, so far either narrative or persuasive. The test could be expanded to include more than a single form or genre. Beginning in Year 5, students could be given a choice from a specified range of forms, as best suits their choice of stimuli, discussed next.

Prompts could be presented in a range of ways, especially if the writing test is 'born digital' and where the test is completed online. A digital placemat would present students with an overall concept or theme and a number of visual and verbal stimuli they would respond to. That should increase the perceived authenticity and relevance of the test to young people and contribute to efforts to arrest the problem, widely reported, of student disengagement from NAPLAN writing.

While the proposal for a digital placemat of stimuli in the context of NAPLAN is new, the use of a range of stimuli is a well-recognised feature in examinations. The writing task that was part of the Core Skills Test in Queensland is a useful reference point for the diverse range of stimuli (for example, artworks, short literary pieces including poetry, news items, reports, and various graphic representations of ideas). Further details can be found at Queensland Curriculum and Assessment Authority (2019).

Informed by the widely reported position that students in Year 3 do not have well-developed keyboarding skills and are

therefore unlikely to produce their best writing online, the online writing test could be introduced in Year 5. This would recognise teachers' insights regarding the classroom priority of mastering handwriting. It would allow opportunity for introducing keyboarding and word processing to strengthen opportunities for students' success in undertaking the online writing test in Year 5.

### **The marking rubrics need simplification**

As explained in Chapter 5, marking of NAPLAN writing is a complex process with 10 traits with reported potential problems of interdependence among them and of a halo effect in applying them.

The marking rubric should be redesigned to include fewer, and conceptually more distinct, traits that would minimise the risk of trait interdependence. The rubric should be consistent with the definition of the writing construct and reflect a balance of the following, consistent with the aspects of writing expected for students at different learning stages:

- higher order authorial skills of audience, ideas and text structure
- the mechanical skills of spelling, punctuation, paragraphing and sentence structure.

The redesigned writing test should take a strengthened focus on language use in context. The language conventions of spelling, grammar and punctuation should be assessed in the writing that students produce, and not separately in decontextualised applications.

The technical issues of scoring including dependencies among criteria, especially in relation to adjacent year levels, should be examined routinely as part of ongoing test evaluation and validity studies. These would open the opportunity for investigating the difficulty and ease with which scorers can separate the criteria for scoring purposes, discriminating among

them and addressing the specified features within each criterion. The most recent National Assessment Program – Literacy and Numeracy (NAPLAN) 2019 Technical Report (June 2020) includes studies of this type and adds to documentation necessary to achieve transparency in test design and implementation and to build confidence at system/sector and school levels and in the wider public.

### **Assessment and reporting could involve teacher judgement**

The redevelopment of the new writing test will take some time. In 2021, students will sit the writing test as already planned. A small trial could be undertaken in 2021 with a selection of students to ensure the proposed redevelopments are fit-for-purpose.

Then, in 2022, the new writing test would be introduced as a sample assessment. This would allow an opportunity to check that the proposed redevelopment delivers a sound approach to the testing of writing.

In 2023, assuming the sample period provides positive evidence, the writing test would revert to census testing.

Noting also the strongly held view of review participants regarding the importance of teachers' professional judgement, it is recommended that the profession plays a key role in the redevelopment of the writing test, including through moderation systems and processes. It is recommended that these be developed initially as part of the sampling methodology in the trial (short-term action), and as a feature of quality assurance systems and processes, with the redeveloped writing test to be re-instated as part of census testing.

The redevelopment work would include a National Calibration Sample, which has been used in NAPLAN to calibrate scales, with a well-structured, random sample of schools to represent each state and territory, taking account of factors such as school size and geolocation with oversampling to ensure

sufficient numbers in some subgroups that have small numbers in the population.

The scripts from the Calibration Sample could be marked centrally as at present but consideration should be given to the use of digital technologies to support national marker training and marker moderation online to bring a stronger national perspective to the marking, which is currently done within states and territories.

The efficiency of the marking could be improved with human judges marking the scripts for authorial aspects of writing (audience, text structure, ideas, persuasive devices/character and setting, cohesion, paragraphing). Automated scoring could be used for scoring selected criteria (spelling, vocabulary, sentence structure and punctuation).

## Recommendation 6

- 6.1 Undertake significant development work on a new writing test to be 'born digital'.
- 6.2 Ensure the new test design demonstrates clear alignment to the Australian Curriculum: English, the Achievement Standards, the General Capabilities, and National Literacy and Numeracy Learning Progressions.
- 6.3 Ensure the new test offers a broadened range of forms including imaginative, persuasive and informative genres, staged across the years of testing (for example, Year 3: imaginative; Year 5: imaginative and persuasive; secondary: persuasive and informative).
- 6.4 Ensure the prompt clarifies to students the audience for the writing.
- 6.5 Extend the assessment time for the writing test to be sufficient for students to be able to draft and edit before producing final copy.
- 6.6 Develop a 'digital placemat' to present students with an overall concept or theme and a number of visual and verbal stimuli they could respond to.
- 6.7 Withdraw the language conventions test as a separate test and assess grammar, punctuation and spelling in the writing test.
- 6.8 Allow Year 3 students to hand write responses and Years 5, 7 and 10 students to write using a computer.
- 6.9 Systematically train students in the use of a keyboard to achieve efficiency before Year 5, with demonstration of fluency in typing to be ongoing throughout schooling.
- 6.10 Simplify the marking rubric with fewer criteria that are conceptually independent.
- 6.11 Trial automated scoring of spelling, vocabulary, sentence structure and punctuation in the writing test, with authorial aspects of writing are to be scored by teachers.
- 6.12 Reinstate the writing test as a census test in 2023, following redevelopment and evidence from the sample trial.
- 6.13 Explore digital approaches to support national marker training and marker moderation online, including the use of exemplars and rich commentaries for a stronger national perspective to the marking and teacher judgement contribution.

## Starting a new time series

With the full implementation of NAPLAN Online, a new time series should be commenced, freed from the constraint of achieving satisfactory horizontal calibration of current scales back to the original 2008

NAPLAN scales. The new scale could be set as the first one was, with a mean of 500 and a standard deviation of 100. Subsequent years' data would then reveal whether and how much means change from this starting point.

### Recommendation 7

Establish a new time series, beginning with the year in which NAPLAN Online is fully implemented.

## Reporting

### Monitoring trends

The 1989 Hobart Declaration committed Australian governments to monitoring national trends in achievement. Since the first 2008 NAPLAN National Report, annual reports have provided national and state and territory analyses of achievement and student progress. They have also considered the differences in achievement among male and female students, and by Indigenous status, language background other than English, geolocation and parental education and occupation. Although some stakeholders would prefer sample rather than census assessments for this purpose, the importance of monitoring trends in achievement over time, across jurisdictions and across equity groups was widely accepted by stakeholders.

In the last decade, NAPLAN results have become a fundamental part of the evidence base for school system accountability and school improvement. Under the current National School Reform Agreement (2018), for example, jurisdictions have agreed to focus on increasing the proportion of students in the top two NAPLAN achievement bands and decreasing the proportion of students in the bottom two achievement bands, including for students in priority equity groups. These commitments cascade down to state and territory budget targets, school system/sector targets and school-level improvement targets. For this reason, annual public reporting of achievement and progress should continue.

### Recommendation 8

Continue to publish annual reports on performance levels – national, state and territory and jurisdiction, as well as for subgroups of interest, such as male and female, Indigenous, students with a language background other than English – and trends in performance levels over time.

## Reporting to schools, parents/carers and the community

There was broad recognition of the value of information for parents/carers provided through individual students' NAPLAN reports. Some parent/carer groups and organisations supporting students with learning difficulties also valued the availability of information independent of the local school context.

Individual students' NAPLAN results currently arrive in schools too late to be useful for guiding individual learning, but the data are often used in triangulation with other standardised tests and teachers' judgements. Fully online testing will reduce the time lag to days, not months, and will increase the value of the assessments in supporting teaching and learning of individual students. Nevertheless, levels of precision of measurements mean that national means are the most precise, school means are more precise for larger schools and measures are least precise for individual students. Appropriately, the ministerial council has recently clarified the importance of teacher judgments and limits to the diagnostic use of NAPLAN.

Many stakeholders were concerned about the impact of public reporting of school-level NAPLAN results. Comparisons among schools in published league tables and on the My School website were widely believed to have made NAPLAN into high stakes tests and led to negative consequences such as teaching to the tests, narrowing of the curriculum and increasing stress for students and teachers. For these reasons, successive ministerial councils have warned against the construction of simplistic league tables. Direct comparisons among statistically similar schools have been removed from the My School website and replaced with comparisons of achievement and progress of students with similar starting points and similar backgrounds. Ministers have also clarified that NAPLAN 'does not measure overall school quality'. There remains, however, a public interest in making available information about each school's contribution to the national effort to make Australia a high equity, high performance nation.

### Recommendation 9

- 9.1 Ministers emphasise that standardised test results should be used in conjunction with school-based assessments in judging students' progress and in reporting to parents/carers.
- 9.2 Ministers emphasise that My School does not compare statistically similar schools but instead provides information on patterns of achievement and growth of similar students from the larger Australian population.

## Ongoing evaluation

The review has recommended substantial changes to the national standardised assessment program. Responding to stakeholders' view that NAPLAN may have narrowed the curriculum, the review has proposed widening the assessment domains beyond literacy and numeracy to include critical and creative thinking in STEM. Concerned that preparation for NAPLAN may have dominated the curriculum in some schools and led to teaching to the test, the review has proposed decreasing the number of assessments and moving the assessments as close as practicable to the beginning of the school year, making the assessments less summative and more formative. Building on feedback about Year 9 students' attitudes to NAPLAN, the review has proposed moving the Year 9 assessment to Year 10, a year in which students are making decisions about their further study options. Based on widespread feedback about the scope, quality and marking of the

current writing assessment, the review has proposed a substantial process of rebuilding take place before writing returns as a census assessment. In preparation for a fully online assessment, the review has proposed that reading and numeracy should be completely redeveloped, capitalising on the flexibility that the digital form offers for content and item development. Finally, in consequence of all these changes to NAPLAN, the review has proposed that a new time series be established, beginning in the year in which NAPLAN Online is fully implemented.

The scope of these proposed changes, if implemented, and the degree to which they succeed in improving the scope and quality of assessment and ameliorating the potential negative consequences of national standardised assessment, are sufficient to warrant a formal evaluation of the program.

### Recommendation 10

Undertake a formal evaluation of any changes made to the national standardised assessment program, with particular attention to the costs and benefits of these changes for students, teachers, schools and school systems/sectors.

## Links to terms of reference and proposed timeline

The ten recommendations above are shown in Table 24 together with their links to the terms of reference for the review and with suggested timing for implementation for each part of each of the recommendations.

**Table 24:** Terms of Reference, recommendations and timeline

Term of Reference	Recommendation	Timeline
<p><b>Term of reference 1</b></p> <p>Determine what the objectives for standardised testing in Australia should be, given its evolution over time – this could be objectives that support:</p> <ul style="list-style-type: none"> <li>• individual student learning achievement and growth</li> <li>• school improvement</li> <li>• system accountability and performance</li> <li>• information for parents/ carers on school and student performance</li> <li>• national, state and territory programs and policies;</li> </ul>	<p><b>Recommendation 1: Standardised assessment and the role of NAPLAN</b></p> <p>1.1 Ministers re-endorse the importance of standardised testing in Australian school education for:</p> <ol style="list-style-type: none"> <li>a. Monitoring progress towards national goals.</li> <li>b. School system accountability and performance.</li> <li>c. School improvement.</li> <li>d. Individual student learning achievement and growth, noting the limitations on use in detailed diagnosis of learning deficiencies and difficulties due to the degree of uncertainty in measures of individual students.</li> <li>e. Information for parents/carers on student and school performance.</li> </ol>	2020

Term of Reference	Recommendation	Timeline
<p><b>Term of reference 2</b></p> <p>Assess how well placed NAPLAN is to meet these objectives, including:</p> <ul style="list-style-type: none"> <li>the appropriateness, accuracy and efficacy of assessment in each domain</li> <li>the effectiveness in tracking student and system progress over time (including the impact of equating, and the placement of the tests in years 3, 5, 7 and 9)</li> <li>alignment with the Australian Curriculum (including any gaps)</li> <li>the impact of the assessment on schools, students and the community.</li> </ul>	<p><b>Recommendation 1: Role of NAPLAN</b></p> <p>1.2 Ministers re-affirm the role of national standardised assessment in fulfilling these purposes.</p> <p>1.3 Continue to conduct national standardised assessment as a census test of student achievement.</p> <p>1.4 Define the purposes and limitations of national standardised assessment by decision of the Ministerial Council in the manner proposed in Table 23 and communicate this on the Australian Curriculum, Assessment and Reporting Authority (ACARA) website and in communications with schools and parents/carers.</p>	<p>2020</p> <p>2020</p> <p>2020</p>
<p><b>Term of reference 3</b></p> <p>Consider the key objectives, uses and features of effective national assessment programs internationally, and how the objectives and performance of NAPLAN compare with this.</p>	<p><b>Recommendation 2: Other national practices</b></p> <p>2.1 Ministers note that national assessment policies and practices vary and that there are no common features of assessment among high-achieving countries.</p>	<p>2020</p>



Term of Reference	Recommendation	Timeline
<p><b>Term of reference 4</b></p> <p>Identify targeted improvements that can be made to standardised testing in Australia in the short-term, including the level of school and student engagement, so it better meets the objectives above.</p>	<p><b>Recommendation 2: Content coverage of tests</b></p>	
<p><b>Term of reference 5</b></p> <p>Identify longer-term objectives, uses and features of standardised testing in Australia within the context of a future national assessment landscape; and consider, in line with these objectives, longer-term improvements that can be made to ensure that Australia has the most efficient and effective system for assessing key literacy and numeracy outcomes at the national level.</p>	<p>2.2 Rename the numeracy test as mathematics, to clarify that it assesses the content and proficiency strands of the Australian Curriculum: Mathematics</p> <p>2.3 Add assessment of critical and creative thinking in science, technology, engineering and mathematics (STEM) to the national standardised census assessment program, except at Year 3, and introduce it only after a period of experimental test development that demonstrates that valid and reliable tests have been developed.</p> <p>2.4 Withdraw the current triennial sample survey of science literacy in Years 6 and 10 and consider replacing it in the triennial cycle with another covering both a subject and a general capability from the Australian Curriculum, such as history and intercultural understanding.</p> <p>2.5 Explicitly map the tests to the National Literacy and Numeracy Learning Progressions to provide insight into student learning progress in the year levels in which the tests are administered.</p> <p>2.6 Jurisdictions investigate students' reasons for absence from NAPLAN testing and seek to reduce the current levels of absence, particularly at the secondary level.</p>	<p>2020</p> <p>2021-22 development 2023 introduction</p> <p>2023</p> <p>2023</p> <p>From 2021</p>

Term of Reference	Recommendation	Timeline
<b>Recommendation 3: Timing of tests</b>		
3.1	Conduct NAPLAN tests as early in the school year as is administratively feasible.	2022
3.2	Set a goal for the results from all NAPLAN Online tests marked online being reported to schools, students and parents/carers within a week of the conclusion of the testing window.	2020
3.3	Continue to administer NAPLAN tests in Years 3, 5 and 7 and replace assessments in Year 9 with assessments in Year 10. Assessments in Year 9 are not to be held in 2021.	2022
<b>Recommendation 4: Rebranding of programs</b>		
4.1	Adopt a new name, Australian National Standardised Assessments (ANSA), in recognition of the changes in the existing tests and the addition of tests of critical and creative thinking in science, technology, engineering and mathematics (STEM).	2022
4.2	Discontinue the National Assessment Program (NAP) sample survey in science literacy with the introduction of ANSA in critical and creative thinking in STEM.	After 2021
4.3	Maintain the National Assessment Program (NAP) sample surveys in civics and citizenship and in information and communication technology literacy on their current three-yearly cycle. Rename the program the National Sample Assessment Program (NSAP).	2022

Term of Reference	Recommendation	Timeline
<b>Recommendation 5: Test development</b>		
5.1	Redevelop the reading and mathematics tests as digital assessments, capitalising on all the flexibility that the digital form offers for content and item form, with no constraint to mirror the current print versions of the tests.	2022
5.2	Develop the new critical and creative thinking in science, technology, engineering and mathematics (STEM) so that it is 'born digital' since it will have no print form that it might have been constrained to match.	2023
5.3	Undertake further development of the branching model and system changes to see if the branching could be based on estimates of achievement derived by the psychometric model and not number of items answered correctly in the testlets.	2022
5.4	Review the level of the National Minimum Standards on the NAPLAN scales to see if they are set too low and progress work on developing additional 'proficient' and 'highly proficient' benchmarks.	2022
<b>Recommendation 6: Writing test changes</b>		
6.1	Undertake significant development work on a new writing test to be 'born digital'.	From 2021
6.2	Ensure the new test design demonstrates clear alignment to the Australian Curriculum: English, the Achievement Standards, the General Capabilities, and National Literacy and Numeracy Learning Progressions.	2022
6.3	Ensure the new test offers a broadened range of forms including imaginative, persuasive and informative genres, staged across the years of testing (for example, Year 3: imaginative; Year 5: imaginative and persuasive; secondary: persuasive and informative).	2022

Term of Reference	Recommendation	Timeline
	6.4 Ensure the prompt clarifies to students the audience for the writing.	2022
	6.5 Extend the assessment time for the writing test to be sufficient for students to be able to draft and edit before producing final copy.	2022
	6.6 Develop a 'digital placemat' to present students with an overall concept or theme and a number of visual and verbal stimuli they could respond to.	2022
	6.7 Withdraw the language conventions test as a separate test and assess grammar, punctuation and spelling in the writing test.	2022
	6.8 Allow Year 3 students to hand write responses and Years 5, 7 and 10 students to write using a computer.	2022
	6.9 Systematically train students in the use of a keyboard to achieve efficiency before Year 5, with demonstration of fluency in typing to be ongoing throughout schooling.	From 2022
	6.10 Simplify the marking rubric with fewer criteria that are conceptually independent.	2022
	6.11 Trial automated scoring of spelling, vocabulary, sentence structure and punctuation in the writing test, while authorial aspects of writing are to be scored by teachers.	2022
	6.12 Reinstate the writing test as a census test in 2023, following redevelopment and evidence from the sample trial.	2023
	6.13 Explore digital approaches to support national marker training and marker moderation online, including the use of exemplars and rich commentaries for a stronger national perspective to the marking and teacher judgement contribution.	2022

Term of Reference	Recommendation	Timeline
	<b>Recommendation 7: New time series</b>	
	Establish a new time series, beginning with the year in which NAPLAN Online is fully implemented.	2022
	<b>Recommendation 8: Annual reports</b>	
	Continue to publish annual reports on performance levels – national, state and territory and jurisdiction, as well as for subgroups of interest, such as male and female, Indigenous, students with a language background other than English – and trends in performance levels over time.	2021
	<b>Recommendation 9: Use with school assessments</b>	
	9.1 Ministers emphasise that standardised test results should be used in conjunction with school-based assessments in judging students' progress and in reporting to parents.	2020
	9.2 Ministers emphasise that My School does not compare statistically similar schools but instead provides information on patterns of achievement and growth of similar students from the larger Australian population.	2020
	<b>Recommendation 10: Evaluation</b>	
	Undertake a formal evaluation of any changes made to the national standardised assessment program, with particular attention to the costs and benefits of these changes for students, teachers, schools and school systems.	2026

## Appendix 1: Summary of recommendations

### Recommendation 1

- 1.1 Ministers re-endorse the importance of standardised testing in Australian school education for:
  - a. Monitoring progress towards national goals
  - b. School system accountability and performance
  - c. School improvement
  - d. Individual student learning achievement and growth, noting the limitations on use in detailed diagnosis of learning deficiencies and difficulties due to the degree of uncertainty in measures of individual students
  - e. Information for parents on student and school performance.
- 1.2 Ministers re-affirm the role of national standardised assessment in fulfilling these purposes.
- 1.3 Continue to conduct national standardised assessment as a census test of student achievement.
- 1.4 Define the purposes and limitations of national standardised assessment by decision of the Ministerial Council in the manner proposed in Table 23 and communicate this on the Australian Curriculum, Assessment and Reporting Authority (ACARA) website and in communications with schools and parents/carers.

### Recommendation 2

- 2.1 Ministers note that national assessment policies and practices vary and that there are no common features of assessment among high-achieving countries.
- 2.2 Rename the numeracy test as mathematics, to clarify that it assesses the content and proficiency strands of the Australian Curriculum: Mathematics.
- 2.3 Add assessment of critical and creative thinking in science, technology, engineering and mathematics (STEM) to the national standardised census assessment program, except at Year 3, and introduce it only after a period of experimental test development that demonstrates that valid and reliable tests have been developed.
- 2.4 Withdraw the current triennial sample survey of science literacy in Years 6 and 10 and consider replacing it in the triennial cycle with another covering both a subject and a general capability from the Australian Curriculum, such as history and intercultural understanding.
- 2.5 Explicitly map the tests to the National Literacy and Numeracy Learning Progressions to provide insight into student learning progress in the year levels in which the tests are administered.
- 2.6 Jurisdictions investigate students' reasons for absence from NAPLAN testing and seek to reduce the current levels of absence, particularly at the secondary level.

### **Recommendation 3**

- 3.1 Conduct NAPLAN tests as early in the school year as is administratively feasible.
- 3.2 Set a goal for the results from all NAPLAN Online tests marked online being reported to schools, students and parents/carers within a week of the conclusion of the testing window.
- 3.3 Continue to administer NAPLAN tests in Years 3, 5 and 7 and replace assessments in Year 9 with assessments in Year 10. Assessments in Year 9 are not to be held in 2021.

### **Recommendation 4**

- 4.1 Adopt a new name, Australian National Standardised Assessments (ANSA), in recognition of the changes in the existing tests and the addition of tests of critical and creative thinking in science, technology, engineering and mathematics (STEM).
- 4.2 Discontinue the National Assessment Program (NAP) sample survey in science literacy with the introduction of ANSA in critical and creative thinking in STEM.
- 4.3 Maintain the National Assessment Program (NAP) sample surveys in civics and citizenship and in information and communication technology literacy on their current three-yearly cycle. Rename the program the National Sample Assessment Program (NSAP).

### **Recommendation 5**

- 5.1 Redevelop the reading and mathematics tests as digital assessments, capitalising on all the flexibility that the digital form offers for content and item form, with no constraint to mirror the current print versions of the tests.
- 5.2 Develop the new critical and creative thinking in science, technology, engineering and mathematics (STEM) so that it is 'born digital' since it will have no print form that it might have been constrained to match.
- 5.3 Undertake further development of the branching model and system changes to see if the branching could be based on estimates of achievement derived by the psychometric model and not number of items answered correctly in the testlets.
- 5.4 Review the level of the National Minimum Standards on the NAPLAN scales to see if they are set too low and progress work on developing additional 'proficient' and 'highly proficient' benchmarks.

## Recommendation 6

- 6.1 Undertake significant development work on a new writing test to be 'born digital'.
- 6.2 Ensure the new test design demonstrates clear alignment to the Australian Curriculum: English, the Achievement Standards, the General Capabilities, and National Literacy and Numeracy Learning Progressions.
- 6.3 Ensure the new test offers a broadened range of forms including imaginative, persuasive and informative genres, staged across the years of testing (for example, Year 3: imaginative; Year 5: imaginative and persuasive; secondary: persuasive and informative).
- 6.4 Ensure the prompt clarifies to students the audience for the writing.
- 6.5 Extend the assessment time for the writing test to be sufficient for students to be able to draft and edit before producing final copy.
- 6.6 Develop a 'digital placemat' to present students with an overall concept or theme and a number of visual and verbal stimuli they could respond to.
- 6.7 Withdraw the language conventions test as a separate test and assess grammar, punctuation and spelling in the writing test.
- 6.8 Allow Year 3 students to hand write responses and Years 5, 7 and 10 students to write using a computer.
- 6.9 Systematically train students in the use of a keyboard to achieve efficiency before Year 5, with demonstration of fluency in typing to be ongoing throughout schooling.
- 6.10 Simplify the marking rubric with fewer criteria that are conceptually independent.
- 6.11 Trial automated scoring of spelling, vocabulary, sentence structure and punctuation in the writing test, while authorial aspects of writing are to be scored by teachers.
- 6.12 Reinstate the writing test as a census test in 2023, following redevelopment and evidence from the sample trial.
- 6.13 Explore digital approaches to support national marker training and marker moderation online, including the use of exemplars and rich commentaries for a stronger national perspective to the marking and teacher judgement contribution.

## Recommendation 7

Establish a new time series, beginning with the year in which NAPLAN Online is fully implemented.



## **Recommendation 8**

Continue to publish annual reports on performance levels – national, state and territory and jurisdiction, as well as for subgroups of interest, such as male and female, Indigenous, students with a language background other than English – and trends in performance levels over time.

## **Recommendation 9**

- 9.1 Ministers emphasise that standardised test results should be used in conjunction with school-based assessments in judging students' progress and in reporting to parents.
- 9.2 Ministers emphasise that My School does not compare statistically similar schools but instead provides information on patterns of achievement and growth of similar students from the larger Australian population.

## **Recommendation 10**

Undertake a formal evaluation of any changes made to the national standardised assessment program, with particular attention to the costs and benefits of these changes for students, teachers, schools and school systems.

## Appendix 2: Review of NAPLAN terms of reference

### Background

NAPLAN has been in place since 2008, and is evolving with the introduction of online testing.

Noting changes in the broader education landscape, both nationally and within states and territories, it's important to consider how NAPLAN can continue to support an effective and contemporary national assessment environment.

The review will be delivered jointly by the state governments of the Australian Capital Territory, Queensland, New South Wales and Victoria.

The review will be informed by and build on work already undertaken or underway, including work that has considered the extent to which NAPLAN has met its original objectives (see below).

### Terms of reference

The review will:

1. determine what the objectives for standardised testing in Australia should be, given its evolution over time - this could be objectives that support:
  - individual student learning achievement and growth
  - school improvement
  - system accountability and performance
  - information for parents on school and student performance
  - national, state and territory programs and policies;
2. assess how well placed NAPLAN is to meet these objectives, including:
  - the appropriateness, accuracy and efficacy of assessment in each domain
  - the effectiveness in tracking student and system progress over time (including the impact of equating, and the placement of the tests in years 3, 5, 7 and 9)
  - alignment with the Australian Curriculum (including any gaps)
  - the impact of the assessment on schools, students and the community;
3. consider the key objectives, uses and features of effective national assessment programs internationally, and how the objectives and performance of NAPLAN compare with this;
4. identify targeted improvements that can be made to standardised testing in Australia in the short-term, including the level of school and student engagement, so it better meets the objectives above;
5. identify longer-term objectives, uses and features of standardised testing in Australia within the context of a future national assessment landscape; and
6. consider, in line with these objectives, longer-term improvements that can be made to ensure that Australia has the most efficient and effective system for assessing key literacy and numeracy outcomes at the national level.

## Other relevant work that the review will need to consider

The review will build on previous and current work including:

- the 2018 Queensland NAPLAN Review
- the 2018/19 Review of NAPLAN Data Presentation
- reviews associated with NAPLAN Online
- other relevant reviews of NAPLAN.

It will not duplicate the outcomes or findings of this work.

The review will also take into account:

- concurrent work on assessment, including commitments in the National School Reform Agreement
- other streams of work which might have implications for assessment goals, including the review of the Melbourne Declaration and updates to the Closing the Gap targets.

## Review process

The review will be led by a panel of up to three members, to be appointed by participating governments. Members will have expertise in assessment, curriculum or other relevant fields. An international expert will be considered as one of the members or as a key advisor.

The review will also be supported by an inter-jurisdictional reference group of practitioners.

Targeted stakeholder consultation (by invitation) will occur in each stage. This will be targeted with the aim of supporting outputs for the reports.

The terms of reference for the review, including the proposed reviewers, will be agreed by relevant ministers.

## Review outputs

### Interim report

Stage 1 will provide an interim report to Education Council in December 2019 with:

- a statement clarifying the objectives of standardised testing in Australia
- suggested immediate improvements to standardised testing in Australia to better meet these objectives
- a summary of longer term issues for investigation that will inform stage 2 of the review.

### Final report

Stage 2 will report to Education Council in June 2020 (\*updated) on a strategic blueprint for standardised testing in Australia, to be considered in concert with the introduction of new assessment approaches (including improvements associated with NAPLAN Online and the national formative assessment capacity).

\*Stage 2 will now report to Education Council in September 2020.

## Appendix 3: List of stakeholders consulted

### Stakeholder consultations

The panel conducted 91 face-to-face and videoconference consultations with 175 individual stakeholders during the review. These organisations and individuals are named below:

Location	Stakeholder institution	Stakeholder name
ACT	ACT Aboriginal and Torres Strait Islander Elected Body	Maurice Walker
	ACT Council of Parents and Citizens Associations	Kirsty McGovern-Hooley
		Veronica Elliot
	ACT Education Directorate	Katy Haire
		Meg Brighton
		Deb Efthymiades
		Robert Gotts
		Kate McMahon
		Mark Huxley
	ACT Government	Simon Tiller
		Ms Yvette Berry MLA
		Joshua Ceramidas
	ACT Principals' Association	Rebecca Hobbs
		Liz Bobos
Association of Independent Schools of the ACT	Wendy Cave	
	Andrew Wrigley	
Australian Education Union – ACT Branch	Joanne Garrison	
	Glenn Fowler	
	Malisa Legyel	
	Sean van der Heide	

Location	Stakeholder institution	Stakeholder name
QLD	Association of Heads of Independent Schools Queensland Branch	Ros Curtis
	Australian Catholic University	Michelle Haynes
	Catholic School Parents Queensland	Carmel Nash
	Catholic Secondary Principals Association of Queensland	Ann Rebgetz
	Independent Schools Queensland	Josephine Wise
		Michael Gilliver
	Joint Council of Queensland Teacher Associations	Danielle Gordon
		Sherryl Saunders
	QLD Aboriginal and Torres Strait Islander Education and Training Advisory Committee	Ned David
		Anita Lee Hong
		Dr Melinda Mann
	QLD Association of Combined Sector Leaders	Brian O'Neill
	QLD Association of Special Education Leaders	Roselynn Anderson
	QLD Association of State School Principals	Leslie Single
	QLD Catholic Education Commission	Dr Lee-Anne Perry
		Yvonne Ries
	QLD Catholic Primary Principals Association	Chris Leeson
	QLD Council of Parents and Citizens' Association	Kevan Goodworth
	QLD Curriculum and Assessment Authority	Chris Rider
		Brian Short
		Tony Cook
	QLD Department of Education	Jim Cousins
		Racquel Gibbons
		Stacie Hansel
		Chris Kinsella
		Amanda O'Hara
Mick O'Leary		
Lesley Robinson		
Robyn Rosengrave		
Pia St Clair		
QLD Government	The Hon. Grace Grace, MP	
QLD Independent Education Union	Dr Adele Schmidt	
QLD Isolated Children's Parents' Association	Tammie Irons	
QLD Secondary Principals' Association	Mark Breckenridge	
QLD Teachers' Union	Cresta Richardson	

Location	Stakeholder institution	Stakeholder name
NSW	Association of Independent Schools NSW	Geoff Newcombe
		Robyn Yates
	Australian Association of Special Education – NSW Chapter	Sally Howell
	Catholic Schools NSW	Dallas McInerney
		Danielle Cronin
	Council of Catholic School Parents NSW/ACT	Peter Grace
	English Teachers Association NSW	Karen Yager
		Mel Dixon
	Family Advocacy	Karen Tippett
	Independent Education Union	Mark Northam
		Pat Devery
	Isolated Children’s Parents Association NSW	Claire Butler
	Lifestart	Sue Becker
	NSW Department of Education	Mark Scott
		Leslie Loble
		Lucy Lu
		Rob Johnston
	NSW Disability Council	Rachael Sowden
	NSW Education Standards Authority	Paul Martin
		Sofia Kesidou
	NSW Federation of Parents and Citizens	Tim Spencer
	NSW Government	The Hon. Sarah Mitchell, MLC
		Meghan Senior
		David Cross
	NSW Maths Association	Karen McDaid
		Maria Quigley
		Darius Samojlowicz
	NSW Parents Council	Teresa Rucinski
	NSW Primary Principals’ Association	Bob Willetts
		Scott Sanford
NSW Teachers Federation	Amber Flohm	
	Denis Fitzgerald	
	Maurie Mulheron	
Professional Teachers Council NSW	David Browne	
Secondary Principals’ Association NSW	Craig Petersen	
	Christine Del Gallo	
SPELD NSW	Georgina Perry	
	Rhonda Filmer	

Location	Stakeholder institution	Stakeholder name
VIC	Australian Education Union (Victorian Branch)	Meredith Peace
		Justin Mullaly
	Catholic Education Melbourne	Bruce Philips
		Simon Lindsay
	Children and Young People with a Disability	Maeve Kennedy
	Council of Professional Teaching Associations of Victoria	Dr Deb Hull
	Department of Education and Training Victoria	Jenny Atta
		Katherine Whetton
		Scott Widmer
		David Howes
		Gabi Burman
		Connie Spinoso
		Robert Mizzi
	Independent Education Union Victoria and Tasmania	Cathy Hickey
	Independent Schools Victoria	Helen Schiele
		Sarah Tielman
	Mathematical Association of Victoria	Peter Saffin
	Parents Victoria	Gail McHardy
		Leanne McCurdy
	SPELD Victoria	Yasotha V
	University of Melbourne	Sandra Milligan
	VIC Association for the Teaching of English	Kate Gillespie
	VIC Association of State Secondary Principals	Sue Bell
	VIC Curriculum and Assessment Authority	Sharyn Donald
		Claude Sgroi
	VIC Government	The Hon. James Merlino MP
		Noah Elrich
		Claudia Laidlaw
	VIC Principals' Association	Anne-Maree Kliman
	VIC Student Representative Council	Joe (Year 11)
	Anna (Year 12)	
	Rielly (Year 11)	
	Sam (Year 10)	
	Astrid (Year 4)	
	Nina (Executive Officer)	

Location	Stakeholder institution	Stakeholder name
National	Australian Council for Educational Research	Geoff Masters
		Catherine McClellan
		Dr. Ray Adams
		Julian Fraillon
	Australian Curriculum, Assessment and Reporting Authority	Peter Titmanis
		Eveline Gebhardt
	Australian Institute for Teaching and School Leadership	Professor John Hattie
		Dr Xian-Zhi Soon
	Australian Literacy Educators Association	Dr Jennifer Rennie
		Dr Jessica Mantei
	Brightpath	Dr Sandy Heldsinger
	Learning Difficulties Australia	Dr Lorraine Hammond
		Dr Molly de Lemos
	Learning Progressions and Online Formative Assessment Initiative	Dr Jenny Donovan
		MultiLit
	Parents for ADHD Advocacy Australia	Dr Robyn Wheldall
		Rimmelle Freedman
	Phil Lambert Consulting	Alex Yourakelis
		Dr Phil Lambert
	Practitioners' Reference Group	Sue Bambling
		Kylie Baxter
		Paul Bennett
		Dale Cain
Gareth Erskine		
Denis Fitzgerald		
Liam Holcombe		
Seir Holley		
Steven Kolber		
Megan Krimmer		
Heidi Livermore		
Isaac Lo		
Julie Ross		
Tonia Smerdon		
Greg Terrell		
Lina Vigliotta	Sophia Williams	
	Dr Pauline Jones	
	Robyn Cox	
	Megan Edwards	
University of Western Australia	Dr Stephen Humphry	
	David Andrich	
UNSW Gonski Institute	Prof. Pasi Sahlberg	
International	NZ Council for Educational Research	Charles Darr
	NZ National Commission for United Nations Educational, Scientific and Cultural Organisation	Robyn Baker
	University of Glasgow	Prof. Louise Hayward



## Appendix 4: International practice in standardised writing assessment

### International testing of writing

A scan of international practices of standardised writing assessment was undertaken to inform this review. The scan included those countries where census testing has been undertaken at the national level, other large-scale standardised assessments of writing implemented through states or provinces, and those countries or states (provinces) that used sampling methodology and opt-in strategies (Table 26).

The scan identified seven countries (Australia, Denmark, Hong Kong, New Zealand, Norway, Singapore, and the United States) that have implemented national large-scale assessments of student writing. Additionally, Canada has implemented census testing in two provinces, Ontario and Manitoba. Australia, Denmark<sup>6</sup>, Hong Kong, and Singapore are identified as the only countries implementing census testing of writing at the national level. Scotland's implementation of standardised testing in 2017 to 2018 has been excluded as Scotland's writing assessment does not include open-ended writing. A further recent testing initiative, the inter-country large-scale assessment of writing known as the Southeast Asia Primary Learning Metrics (SEA-PLM), has been included due to the distinctive nature of the assessment.

The scan took as its focus the following – the purpose of the writing assessment; the methodology as either census or sample testing; the forms of writing selected and the scope of the testing (extended writing, spelling, conventions, punctuation); the prompts and modes of assessment; the test conditions for completing the assessment, including time duration and completion online or handwritten; the function of criteria scoring; scoring (human judgement and machine marking); moderation as part of quality assurance processes; and reporting of results from the writing tests (see Figure 12). These feature areas relevant to a priority area that emerged during this review of NAPLAN, namely the role of the profession in standardised testing of writing.

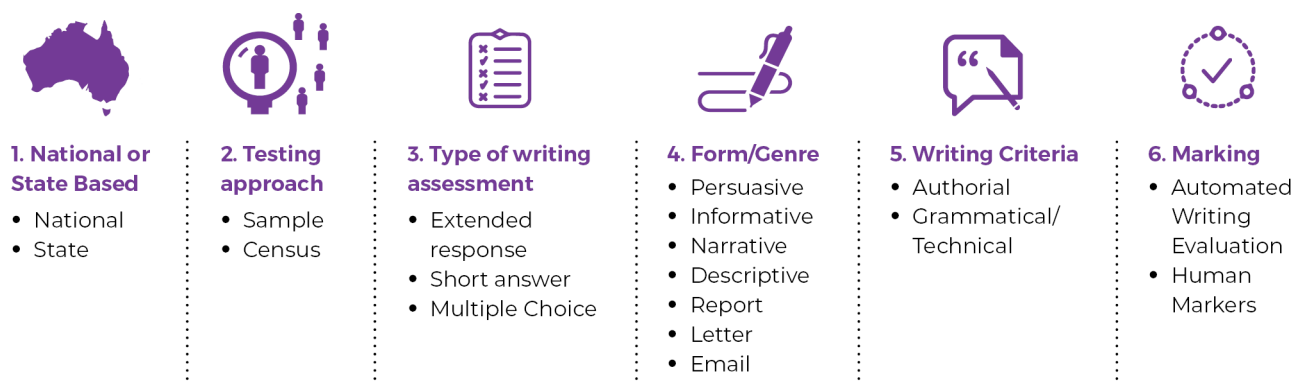
### Australia

Australia's National Assessment Program – Literacy and Numeracy (NAPLAN) is an annual assessment for students in Years 3, 5, 7 and 9, that 'tests the fundamental disciplines of literacy and numeracy' (ACARA, 2017, p. 1). NAPLAN tests skills in the following four areas (or 'domains') – reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. NAPLAN has been undertaken nationally since 2008. The NAPLAN writing assessment aligns with the Australian English curriculum and includes the types of texts that are essential for students to master if they are to be 'successful learners, confident and creative individuals, and active and informed citizens' (MYCEETYA, 2008a, p. 7). The underlying construct of NAPLAN writing assessment tasks is independent of year level;

---

<sup>6</sup> The census testing of writing in Denmark is undertaken in public schools only with schools in other sectors able to opt-in.

NAPLAN prompts span two-year levels, Years 3 and 5, and Years 7 and 9 respectively; the design of NAPLAN marking rubrics applies to all prompts, irrespective of year level, with the reporting of the results on a single scale for all students in Years 3 to 9. There are ten criteria (ACARA, 2017) that students' writing is assessed against, divided into genre-based or authorial criteria and technical or grammatical criteria.



**Figure 12:** Categories of validity evidence in large-scale assessment of writing

The assessment of writing focuses on two genres (Persuasive and Narrative). One is selected each year for assessment with two prompts chosen, one for students in Years 3 and 5 and another chosen for students in Years 7 and 9. Marking is conducted externally with current and ex-teachers, but the writing assessment is marked independently across the six states and two territories, with centres monitored by the appointed members or the Marker Quality Team. Teachers have a limited role in the system; they do not contribute to selecting assessment writing prompts or participate in marking the student's scripts unless they apply to mark NAPLAN external to their teaching roles. Results from student writing are provided to all schools, students and parents/carers, although there is limited understanding of how much the data are used for informing teaching and learning (Hardy, 2014).

While the intent of Australia's NAPLAN is not to determine the pathway for secondary schooling, the testing in Years 3, 5, 7 and 9 is part of a broader goal in a 'national approach to setting educational expectations' and to provide 'a national consistent measure to determine whether or not students are meeting important educational outcomes' (ACARA, 2020a, unpaginated).

In Australia, school results are currently published online on the platform My School (ACARA, 2020c). This platform allows comparison of schools, with reported concerns including the resultant narrowing of the curriculum (Comber, 2012), scrutiny of principals and teachers (Hardy, 2014, Thompson, 2013), and judgement from media and the community of school performance (Dulfer et al., 2012, Gorur, 2016).

### Canada: Ontario

The Canadian province of Ontario conducts yearly, large-scale, census testing in Years 3, 6, 9 and 10 in reading, writing and maths. Assessment of writing occurs in Year 3 (primary), and Year 6 (junior) and Year 10. Students in Years 3 and 6 complete the Education Quality and Accountability Office (EQAO) elementary assessment that tests students with extended response questions. These are marked against a holistic criterion that focuses on topic development and conventions. EQAO results are reported at the provincial, school board and school levels and are used by the Ministry of Education, district school boards and schools to improve learning, teaching and student achievement (EQAO, 2007).

Year 10 students complete the Ontario Secondary School Literacy Test (OSSLT), which is a provincial test of literacy (reading and writing) skills students have acquired by Grade 10. It is based on the literacy skills expected in the Ontario Curriculum across all subject areas up to the end of Grade 9. The writing assessment involves two extended writing tasks and two short-writing tasks (six lines each). EQAO reports on student achievement at the individual, school, board and provincial levels. Students who participate in the OSSLT receive an Individual Student Report that indicates whether they have successfully completed the OSSLT. Schools and boards will also receive a report that provides aggregated achievement results, aggregated contextual data about students' literacy preferences and practices and provincial results (EQAO, 2020).

EQAO recruits as many teacher-markers (that is, members of the Ontario College of Teachers) as possible and fills the complement with retired educators and qualified non-educators (defined as other-degree markers). All potential scorers must pass a qualifying test to ensure they have sufficient proficiency in English or French. A blind scoring model is used with two markers scoring the scripts.

If the two scores are in exact agreement, that score is assigned to the student. If the two scores are adjacent, the higher score (for reading and short-writing tasks) or the average of the two scores (for news reports and paragraphs expressing an opinion) is assigned to the student. If the two scores are non-adjacent, the response is scored again by an expert scorer to determine the correct score for the student (EQAO, 2020, p. 15).

### **Canada: Manitoba**

The Manitoba Education Department takes the position that the primary role of assessment is to 'enhance teaching and improve student learning and supports this through the Provincial Assessment Initiative and the Provincial Assessment Program' (Manitoba, 2020a, unpaginated). The primary purpose of the Middle Years Assessment policy is to enhance student learning and engagement through classroom-based assessment processes that build student awareness and confidence in learning.

Manitoba students in Grades 8 undergo classroom-based assessments in writing. Teachers base assessments of their students on their observations, conversations with students, and their evaluations of students' classroom-based work. They report on student performance levels as of the last two weeks of January. Evaluation criteria, including the competencies and scoring scales with descriptions and examples, are provided by the department and are used by teachers when reporting achievement results for these assessments to parents/carers and to the department (Manitoba Education Department, 2020b). A summary report is published on the department's website as aggregated results (Manitoba Education and Training, 2020c)

While there is a national test in Canada called the Pan-Canadian Assessment Program, it is a sample test and does not assess the domain of writing.

## Denmark

Danish National Tests were implemented in the public compulsory schools as a mean of evaluating the performance of the public-school system (Folkeskole) in 2010. Census assessment of Year 9 Folkeskole students is not compulsory for students in private schools and these schools may opt-in. Denmark's extensive test program consists of ten mandatory tests in six subjects in grades 2 through to grade 8, although the assessment of writing is not included in these national tests.

At the conclusion of Years 9 and 10, Folkeskole students complete school-leaving examinations, which are compulsory in Year 9 but voluntary in Year 10. Of the compulsory five examinations, one is a written examination in Danish. These assessment genres can take the form of literary fiction, journalistic genres or an essay. Text materials form a part of the assessment prompt to provide inspiration for student writing (Krogh, 2018). Examinations are graded by the classroom teacher and an external censor on a 7-point ordinal scale: -3, 00, 02, 4, 7, 10, and 12, with marks 02 or above considered a pass (Beuchert & Nandrup, 2015).

In conjunction with the leaving examinations, a mandatory interdisciplinary project assessment needs to be submitted and is assessed 'in a written statement on the content, working process and presentation of the final result' (Ministry of Children and Education, 2020, unpaginated). At the student's request, a mark may be awarded and be indicated in the leaving certificate.

According to an OECD Review of the Evaluation and Assessment in Education (2011), the introduction of the national tests offered monitoring information on the Folkeskole at different stages in compulsory education and provided 'the first real opportunity to reliably monitor progress in educational outcomes over time against the national Common Objectives. However, the lack of inclusion of the private sector limits their national monitoring value' (Shewbridge, Jang, Matthews, & Santiago, 2011 p.9). In this review it was also noted that there is a need for the national tests to include open-ended questions in Years 2 to 8.

During the lower secondary leaving examination students are provided "small booklets containing assignments and various text materials are made available to students" (Krogh, 2018, p. 10) as support material for the completion of the writing assessment. Visuals are typically embedded with the prompt and provided as 'inspiration for student writing, only rarely to be addressed explicitly' (Krogh, 2018, p. 10).

Marking is conducted by teachers and 'a sample of examinations are marked by an external censor [who] provides an equitable way to judge whether students have achieved the national Common Objectives' (Shewbridge et al., 2011, p. 51). The secondary leaving examination is aligned to the National Common Objectives and four reports are customised for different stakeholders. Students, parents/carers and teachers are provided with student's test results, school leaders are provided with their school's test results, the Municipality receives the average score of the schools in the municipality, and at the national level, the national average test result for all schools is published and made available to the public (Houlberg et al., 2016).

## Hong Kong

The Territory-wide System Assessment (TSA) is administered by the Hong Kong Examination and Assessment Authority. The intent of the assessment is to provide schools with information about the performance of students in primary 3 (P.3), primary 6 (P.6) and secondary 3 (S.3) including strengths and weaknesses against specific Basic Competencies and to 'to help schools understand students' overall academic standards in the main key learning areas and as a reference for the follow-up action of learning and teaching' (Hong Kong Examinations and Assessment Authority, 2019, p. 1). A related intent is to 'help the Government to review policies and to provide focused support to schools' (Hong Kong Examinations and Assessment Authority, 2020, unpaginated). The TSA began as a national census test for students in Years 3, 6 and 9, however starting from 2012, an alternate-year arrangement has been adopted in the Year 6 assessment, with Year 3 students assessed using a sampling method. The assessment is low-stakes and does not determine secondary placements or pathways. The writing assessment design focuses on letter, narrative or descriptive writing with the criteria for Years 3 and 6 focusing on Content and Language with Year 9 including two extra criteria, Organisation and Features.

Support is offered to teachers through an online platform (web-based learning and teaching support) which provides teaching activities and materials for addressing students' 'relevant learning difficulty in Basic Competencies' (Hong Kong Examinations and Assessment Authority, 2019, p. 1). Schools are also encouraged to make use of the data to adjust teaching plans and teaching strategies.

Moderation Committees are formed, which consist of serving teachers or school heads, a professional staff member of a tertiary institute, and subject officers and managers from the Education Bureau. Students receive randomly allocated writing assessment prompts with teachers playing no role in prompt selection. Markers for the TSA are all qualified serving teachers with a requirement of an attainment of the 'Language Proficiency Assessment for Teachers in English' (Hong Kong Examinations and Assessment Authority, 2019) before being employed as a marker. Extensive training precedes the scoring of the writing assessments.

### The school reports

provide detailed data on the performance in the sub-papers for individual learning dimensions (skills) of individual subjects as well as data at the territory-wide level for reference to help schools identify the overall strengths and weaknesses of students in learning (Hong Kong Examinations and Assessment Authority, 2019, p. 15).

The intent is for schools to use 'the relevant data to adjust their school-based curriculum, teaching strategies and activities' (p. 15). The performance of individual students is not included in all reports, 'which are strictly confidential and provided for schools' reference only' (p. 15).

## New Zealand

The New Zealand National Monitoring Study of Student Achievement (NMSSA) is a national sample test that assesses eight learning areas specified in the New Zealand Curriculum. One test of writing was conducted in 2012, which specifically targeted Years 4 and 8 students. The NMSSA aligned to the Literacy Learning Progressions and English Curriculum, and while the NMSSA only tested writing in 2012, a National Report was published. The mode of delivery was pencil and paper and the task focused on the narrative genre. The writing assessment was based on the Electronic Assessment Tools for Teaching and Learning (e-asTTle) framework with external markers trained to rate student writing.

The testing of writing as part of NMSSA has not been undertaken since 2012. Instead the New Zealand Ministry of Education provides an optional assessment for students called e-asTTle, an online assessment tool developed to assess students' achievement and progress in reading, mathematics, writing, and in pānui, pāngarau, and tuhituhi. The writing tool has been developed to assess students in Years 1 to 10 and can be used formatively at any time during the year as determined by the teacher. Further, teachers mark the writing assessments with support from published manuals and the data is used as part of the teachers' wider collection of information about students' learning and progress.

New Zealand provides e-asTTle free of cost to all schools. The flexibility of the online assessment provides autonomy for teachers, enabling them to make decisions regarding timing of the assessment, selection of appropriate writing prompts for students, and engages their professional judgement in scoring student scripts through the provision of support material with manuals and exemplars to mark student writing. E-asTTLE is aligned to the New Zealand Curriculum Reading and Writing Standards for Years 1 to 8. The results are reported through the platform and professional support is provided to teachers via extensive and relevant resources for addressing and raising student achievement. These resources are also available to parents/carers, school managers and trustees (e-asTTle, 2020).

## Norway

National census tests in Norway were launched initially in 2005 in response to concerns that students were not receiving adequate instruction in 'key competencies' (Official Norwegian Report (green paper): NOU 2002), however an evaluation of the writing test found low rater reliability resulting in the discontinuation of the writing test in 2006 (Skar, 2017). Instead, the Norwegian Directorate for Education and Training commissioned the National Writing Centre (NWS) to develop the Norwegian Sample-Based Writing Test (NSBWT), an annual writing test on a nationally representative sample of students in primary and lower secondary school. The writing assessment design created the option of three genres, persuasion, description and imagination and was marked against five rating scales. A sample of teachers was also asked to complete a survey relating to student writing as part of the report of the writing sample results.

The intent of the project was to 'set up a national panel of raters (NPR), consisting of teachers, with the purpose of 1) establishing a strong interpretive community and 2) having in place a panel that would reliably rate the NSBWT' (Skar & Jølle, 2017, p. 1). The tasks and rating scales of the NSBWT were represented in the theoretical model called the wheel of writing, a theoretical frame for specifying standards of writing proficiency for the teaching profession (Berge, Evensen, & Theygesen, 2016). The NSBWT defined writing proficiency as 'the proficiency to engage in an act of writing using necessary mediating tools' (Skar, 2017, p. 5). Fulfilling the intent of the project was to work with teachers as raters and also to have teachers complete survey questions relating to – the relevance of the writing assessment task; student motivation to complete the task; and sufficient time for task completion (Skar, 2017). Results from marking the sample of student scripts and data from the survey, were published in a National Technical Report.

While the Norwegian Centre for Writing Education and Research (The Writing Centre) is still in operation to support teacher learning in writing education, the NSBWT was discontinued in 2017 due to state budget cuts (Jeffery, Elf, Skar & Campbell Wilcox, 2019).

## **Singapore**

Singapore's Primary School Leaving Examination (PSLE) is a national census, high-stakes, large-scale assessment that is used to 'gauge students' readiness and aptitude to proceed to higher level of schooling – either to select or place students appropriately' (SEAMEO INNOTECH, 2015, p. 133). The PSLE writing assessment is divided into two types of writing defined as 'situational' and 'continuous'. The situational writing requires students to write a short functional piece that could be either a letter, email or report. The continuous writing requires students to write an extended response on a given topic and multi-modal options may be used as prompts. Writing tests span multiple languages inclusive of, English and other mother tongue languages for example, Bengali, Gujarati, Hindi, Panjabi and Urdu. The writing assessment is designed by 'a professional panel of specialists with assessment and subject expertise' (Singapore Examinations and Assessment Board, 2020, unpaginated) and is aligned to the Primary School Curriculum.

The results from the PSLE contribute to decisions about a students' secondary school pathway. On completion of the tests, and dependent on the results, students will progress along one of three options – an express course, a normal academic course, or a normal technical course (SEAMEO INNOTECH, 2015). The PSLE aligns with what Verger, Parcerisa and Fondevila (2019) define as one of three uses or purposes of national assessments namely the 'Assessment for students' certification, streaming and selection purposes [and are] standardised examinations that are high stakes for students, but not necessarily for schools' (p. 10). As Singapore's PSLE is designed for streaming and selection purposes for secondary pathways, the PSLE could be seen as high-stakes for students.

Key Marking Personnel (KMP) made up of Principals, Vice-Principals and Heads of Department engage in moderation practices of the marking scheme and application of the scores to the scripts. Once consensus has been reached regarding the application of criteria to writing scripts by the KMP, markers begin the process of scoring student scripts. KMP at various levels sample check the marking to ensure the accuracy and consistent application of the criteria to the student scripts.

The Ministry of Education and the Singapore Examinations and Assessment Board present a joint press release on student results that indicates the total number of primary students involved as well as percentages of students progressing into the Express course, the Normal (Academic) course, and the Normal (Technical) course.

### **United States of America**

In the United States, the National Assessment of Educational Progress or NAEP is the 'largest nationally representative and continuing assessment' (National Center for Education Statistics (NCES), 2019, p. 1) of student achievement in civics, economics, geography, mathematics, music and visual arts, reading, science, technology and engineering literacy, U.S. history, and writing. The assessment is 'a congressionally mandated project administered by the NCES within the U.S. Department of Education and the Institute of Education Sciences (IES)' (NCES, 2019, p. 2). This assessment involves a representative sample of students across the country in the domain of writing with the test including Years 8 and 12 students nationally in 2011. Further testing in 2017 of students in Years 4 and 8, was conducted online and the report targeted for release in 2020 expected to provide insight into the future design and administration of digitally based NAEP writing assessments (NCES, 2017). The next NAEP writing assessment is scheduled for 2029 (NCES, 2020a). The NAEP is aligned to the Common Core State Standards.

The nature of the writing assessment is the choice of two types of genres, that is, to explain or to persuade. The assessment is online, and prompts are presented to students in a variety of ways including text, audio, photographs, video or animation. Standard tools for editing, formatting and viewing writing are included as part of the online platform as is the ability to use spell check and a thesaurus. Marking is centralised and 25% of scripts are double-marked and 5% are check marked (NCES, 2020b). Results from the NAEP are published in the form of a National report.

In addition to the US NAEP testing, individual states may opt to implement state-based tests of writing. As an example, two consortia that test writing across US states are the Smarter Balanced Summative consortium and the Partnerships for Assessment of Readiness for College and Careers (PARCC) consortium. Both provide writing tests to member states across multiple stages of schooling and multiple text types. The tests are designed to function as summative assessments of student writing skills, linking to the US Common Core State Standards, and can be implemented annually according to each state's testing window.

Another example of state-based implementation of large-scale assessment of writing in the US is the New York State English Language Arts tests (ELA). This is an annual census test for all students in Years 3 to 8. The writing test design asks students to complete both short-response writing and an extended response. Extended response questions are designed to assess writing from sources and will ask students to express a position and support it with textual evidence to support their ideas. The intent of ELA is 'to ensure that schools prepare students with the knowledge and skills they need to succeed in college and in their careers' (New York State Education Department, 2020, p. 1). Results from the assessment are used to assess whether students are meeting the New York State P-12 Learning Standards.



### **Inter-country large-scale standardised assessment**

Lastly, in 2012, the Southeast Asian Ministers of Education Association (SEAMEO) and UNICEF initiated the Southeast Asia Primary Learning Metrics (SEA-PLM) in an effort to assess and monitor students' acquisition of knowledge and skills and to further improve the quality of primary education in Southeast Asia. This distinctive inter-country, large-scale assessment of writing aims to inform policy makers in the participating countries of the progress of educational development in their respective countries. SEA-PLM assesses mathematical literacy, reading literacy, writing literacy and global citizenship.

Of the eleven countries that are involved in the program, only six countries (Myanmar, Vietnam, Lao PDR, Cambodia, Malaysia, Philippines) implement the writing literacy assessment component of the suite of assessments. The sample-based assessment targets Year 5 students and utilises the genres of narrative, description, persuasion, instructional and transactional as part of the assessment. The program is designed as a sample-based assessment, with the intent of generating information to assist other stakeholders, such as teachers, parents/carers and students, in improving learning at the local level (UNICEF & SEAMEO, 2019a).

The assessment reflects a global policy priority of increasing access to education for children and the use of data to monitor progress towards national targets for improvement (UNESCO, 2019). As part of SEA-PLM, the cross-national sample writing assessment 'marks the first cross-national initiative to measure writing literacy understood as the ability to construct meaning by generating a range of written texts to express oneself and communicate with others to meet personal, societal, economic and civic needs.' (UNESCO, 2019, p. 42). The SEA-PLM draws on the curricula of six countries (mentioned above) to develop the assessment framework.

The extended writing assessment criteria had the challenge of measuring writing in a multilingual assessment and achieving equivalence across languages. To work through this challenge,

'the SEA-PLM writing literacy assessment model treats some writing processes as common across languages, while others may be treated as applicable only to one language or to a group of languages. This approach will yield some comparisons between writing performance in different languages, while recognising the particular characteristics of individual languages' (UNICEF & SEAMEO, 2019a, p. 41).

The Australian Council for Educational Research (ACER) was contracted to design and implement the first round of SEA-PLM assessment. How students perform and the success of the testing is yet to be released, with the results from the first SEA-PLM due in 2020. The criteria used are in conjunction with numeric scores, but a degree of flexibility needed to be created to achieve equivalence across languages. An example of the model for writing assessment processes by language and text type is shown in Table 25 (UNICEF & SEAMEO, 2019a, p. 41).

**Table 25:** Model for assessment of writing in multiple languages

Process	Application by language	Application by type of text
Generate ideas	Apply across languages	Vary by text type
Control structure	Apply across languages	Vary by text type
Manage coherence	Apply across languages	Apply cross text types
Use vocabulary	Apply across languages	Apply cross text types
Control syntax and grammar	May vary by language	Apply cross text types
Other language-specific features	May vary by language	Apply cross text types

Data standards have been implemented to ensure the ‘comparability of data across each of the participating countries, the responses from all test participants should be coded following a single coding scheme’ (UNICEF & SEAMEO, 2019b, p. 21) and ‘coders’ are recruited and trained to adhere to ‘agreed procedures’. While the intent is for the results to be used by policy makers and to help teachers inform practice, the role of the teacher in the system appears to be removed and replaced by external contractual arrangements. Subject to conditions of implementation, expectations of data use, and the engagement of the profession, the role of teachers has the potential to grow.

## Summary

This section has described international practice in the assessment of writing in seven countries. Of these, Australia, Hong Kong, Denmark and Singapore are selected countries currently implementing national census testing of writing.

How a country tests writing reflects interrelated decisions about – the purposes of testing; the stages of schooling to be included; the curriculum domains and related forms of writing to be tested; how the writing is to be scored (including criteria, judgement method, human and machine scoring); the role of the profession; quality assurance processes including online moderation; intended uses of the reported results, and to whom and how they are released. All these matters are central to a decision about whether a test is fit-for-purpose. Why countries implement or do not implement national census testing of writing and their approach to implementation merit further exploration. Finally, the scan shows growing interest internationally in the role of the profession in national testing, including for system monitoring purposes, and the role of teacher judgement in scoring and interpreting the results for use in the classroom.

**Table 26:** International large-scale assessment of writing

Country	National or state-based	Sample or Census	Stage or schooling	Writing genre	Criteria	Technology	Marking
<b>Australia:</b> National Assessment Program Literacy and Numeracy (NAPLAN)	National	Census	3,5,7,9	<ul style="list-style-type: none"> <li>• Persuasion</li> <li>• Narrative</li> </ul>	<ul style="list-style-type: none"> <li>• Audience</li> <li>• Text Structure</li> <li>• Ideas</li> <li>• Persuasive devices/ Character and setting</li> <li>• Vocabulary</li> <li>• Cohesion</li> <li>• Paragraphing</li> <li>• Sentence structure</li> <li>• Punctuation</li> <li>• Spelling</li> </ul>	Paper and pencil and online	Human markers
<b>Canada:</b> Ontario  EQAO Elementary Assessments	State	Census	3,6	<p><b>Year 3</b></p> <ul style="list-style-type: none"> <li>• Personal opinion</li> </ul> <p><b>Year 6</b></p> <ul style="list-style-type: none"> <li>• Letter</li> </ul>	<ul style="list-style-type: none"> <li>• Topic development [0-4]</li> <li>• Conventions [0-3]</li> </ul>	Paper	Human markers
<b>Canada:</b> Ontario  OSSLT	State	Census	10	<ul style="list-style-type: none"> <li>• News report (one page)</li> <li>• Series of paragraphs expressing an opinion (two pages)</li> <li>• Two short-writing tasks (six lines each).</li> </ul>	<ul style="list-style-type: none"> <li>• Clarity of communication,</li> <li>• Development of ideas,</li> <li>• Organisation</li> <li>• Language conventions and usage.</li> </ul>	Paper	Human markers

Country	National or state-based	Sample or Census	Stage or schooling	Writing genre	Criteria	Technology	Marking
<b>Canada:</b> Manitoba	State	Census	8	<ul style="list-style-type: none"> <li>Expository</li> <li>Classroom writing summative results based on achievement as of the last two weeks of January.</li> </ul>	<ul style="list-style-type: none"> <li>Ideas (generates, selects and organises)</li> <li>Language (word choice and sentence patterns)</li> <li>Conventions (spelling, grammar, and/or punctuation)</li> <li>Resources (spell-checker, thesaurus, dictionaries) to edit and proofread.</li> </ul>	Online (using spell check, thesauruses, and dictionaries) to edit and proofread.	Teachers mark own classroom writing and submit scores
<b>Denmark:</b> School-leaving examinations (SLE)	National (for Folkeskole schools)	Census	9, 10	<p>Two types of written assessment:</p> <ol style="list-style-type: none"> <li>Written examination under exam conditions <ul style="list-style-type: none"> <li>Literary fiction</li> <li>Journalistic genres</li> <li>Essay</li> </ul> </li> </ol> <p>Use of text materials, typically embedded in visuals.</p> <ol style="list-style-type: none"> <li>A mandatory project assignment gives students the opportunity to complete and present an interdisciplinary project. The project assignment is assessed in a written statement on the content, working process and presentation of the final result.</li> </ol>	<p>Marks are awarded according to a 7-point marking scale</p> <p>12 Excellent performance, high command, few minor weaknesses</p> <p>10 Very good performance, high command, with minor weaknesses</p> <p>7 Good performance, good command, some weaknesses</p> <p>4 Fair performance, some command, major weaknesses</p> <p>2 Meeting only the minimum requirements for acceptance</p> <p>0 Does not meet the minimum requirements for acceptance</p> <p>-3 Unacceptable in all respects</p>	Paper	Teacher and external 'censor'

Country	National or state-based	Sample or Census	Stage or schooling	Writing genre	Criteria	Technology	Marking
<b>Hong Kong:</b> Territory-wide System Assessment (TSA)	National (with the exemption of Year 3)	Census	3, 6, 9	<ul style="list-style-type: none"> <li>• Letter</li> <li>• Narrative</li> <li>• Description</li> </ul>	<p><b>Year 3 and 6</b></p> <ul style="list-style-type: none"> <li>• Content (level of detail, structure, ideas and clarity).</li> <li>• language (e.g. vocabulary, sentence patterns, cohesive devices, grammar, punctuation, capitalisation and spelling)</li> </ul> <p><b>Secondary (Year 9) has two additional criteria</b></p> <ul style="list-style-type: none"> <li>• Organisation (Paragraphs, coherent links and connectives)</li> <li>• Features (Structure e.g. letter format, description and speech in narration).</li> </ul>	Paper and pencil	Human markers
<b>Myanmar, Vietnam, Lao PDR, Cambodia, Malaysia, Philippines:</b> Southeast Asia Primary Learning Metrics (SEA-PLM)	Inter-Country	Sample	5	<ul style="list-style-type: none"> <li>• Narrative</li> <li>• Descriptive</li> <li>• Persuasive</li> <li>• Instructional</li> <li>• Transactional</li> </ul>	<ul style="list-style-type: none"> <li>• Generate ideas</li> <li>• Control structure</li> <li>• Manage coherence</li> <li>• Use vocabulary</li> <li>• Control syntax and grammar</li> <li>• Other language-specific features (spelling, character formation and punctuation)</li> </ul>	Paper and pencil	Human markers
<b>New Zealand:</b> -e-asTTle	National	Optional	1-10	<ul style="list-style-type: none"> <li>• Describe</li> <li>• Explain</li> <li>• Recount</li> <li>• Narrate</li> <li>• Persuade</li> </ul>	<ul style="list-style-type: none"> <li>• Ideas</li> <li>• Structure and language</li> <li>• Organisation</li> <li>• Vocabulary</li> <li>• Sentence structure</li> <li>• Punctuation</li> <li>• Spelling</li> </ul>	Paper and Pencil. Scripts are scored offline and marks entered into e-asTTle system.	Human markers (classroom teachers)

Country	National or state-based	Sample or Census	Stage or schooling	Writing genre	Criteria	Technology	Marking
<b>New Zealand:</b> National Monitoring Study of Student Achievement (NMSSA)	National	Sample	4,8	<ul style="list-style-type: none"> <li>Narrative</li> </ul>	<p><b>Writing for a variety of purposes</b> Based on the e-asTTle framework (see above).</p> <p><b>Process of writing</b> This comprised seven elements:</p> <ul style="list-style-type: none"> <li>Audience awareness</li> <li>Planning</li> <li>Crafting/writing</li> <li>Revising and editing</li> <li>Proofreading</li> <li>Feedback</li> <li>Publishing</li> </ul>	Paper and pencil, one to one interviews and questionnaire.	Human markers
<b>Norway:</b> Norwegian Sample-Based Writing Test (NSBWT) 2010-2016	National	Sample	5, 8	<ul style="list-style-type: none"> <li>Persuade</li> <li>Describe</li> <li>Imagine</li> </ul>	<ul style="list-style-type: none"> <li>Writer-reader interaction</li> <li>Content</li> <li>Text structure</li> <li>Language use</li> <li>Coding competencies (e.g. grammar, spelling and punctuation)</li> </ul>	Paper and pencil	Human markers
<b>Singapore:</b> Primary School Leaving Certificate (PSLE)	National	Census	6	<p>Situational Writing:</p> <ul style="list-style-type: none"> <li>letter</li> <li>email</li> <li>report</li> </ul> <p>Continuous Writing:</p> <ul style="list-style-type: none"> <li>Three pictures will be provided on the topic offering different angles of interpretation.</li> <li>Candidates may also come up with their own interpretation of the topic.</li> </ul>	<ul style="list-style-type: none"> <li>AO1 write to suit purpose, audience and context in a way that is clear and effective</li> <li>AO2 use appropriate register and tone in a variety of texts</li> <li>AO3 generate and select relevant ideas, organising and expressing them in a coherent and cohesive manner</li> <li>AO4 use correct grammar, spelling and punctuation</li> <li>AO5 use a variety of vocabulary appropriately, with clarity and precision</li> </ul>	Paper and pencil	Human markers

Country	National or state-based	Sample or Census	Stage or schooling	Writing genre	Criteria	Technology	Marking
<b>United States:</b> PARCC (Partnerships for Assessment of Readiness for College and Careers)	State	Optional	K-11	<ul style="list-style-type: none"> <li>• Research Simulation Task (RST)</li> <li>• Literacy Analysis Task (LAT)</li> <li>• Narrative Writing Task (NWT)</li> </ul>	<ul style="list-style-type: none"> <li>• Development of ideas</li> <li>• Organisation</li> <li>• Clarity of language</li> <li>• Knowledge of language and conventions.</li> </ul>	Online	External human markers
<b>United States:</b> New York English Language Arts (ELA) test	State	Census	3-8	<ul style="list-style-type: none"> <li>• Text provided:</li> <li>• Short answers</li> <li>• Text provided: Extended Response</li> </ul>	<p><b>Short response (2 point)</b> <b>Holistic Rubric</b></p> <ul style="list-style-type: none"> <li>• Make a claim,</li> <li>• Take a position or</li> <li>• Draw a conclusion, (complete sentences)</li> </ul> <p><b>Extended response (4 point)</b> <b>Holistic Rubric</b></p> <ul style="list-style-type: none"> <li>• Content and Analysis,</li> <li>• Command of Evidence</li> <li>• Coherence, Organisation and Style</li> <li>• Control of conventions</li> </ul>	Option for computer or pen/paper	Human markers
<b>United States:</b> NAEP 2017 US national sample assessment	National	Sample	8,12	<ul style="list-style-type: none"> <li>• Explanation</li> <li>• Persuasion</li> <li>• Convey experience (real or imagined).</li> </ul>	<ul style="list-style-type: none"> <li>• Development of idea</li> <li>• Organisation of ideas</li> <li>• Language facility and conventions</li> </ul>	Online (Tasks): <ul style="list-style-type: none"> <li>• Text</li> <li>• Audio</li> <li>• Photographs</li> <li>• Video</li> <li>• Animation.</li> </ul>	Human markers

# References

Academic Assessment Services. (2020). *Tracking student progress*. Accessed on 29 June 2020 on <http://www.academicassessment.com.au>.

Australian Bureau of Statistics. (2020). *2001 Census of Population and Housing – Geographic Areas*. Accessed on 27 June 2020 on [https://www.abs.gov.au/websitedbs/d3110124.nsf/497f562f857fcc30ca256eb00001b48e/53bbe9630b24d6f4ca256c3a000475b8!OpenDocument#Collection%2520District%2520\(CD\)](https://www.abs.gov.au/websitedbs/d3110124.nsf/497f562f857fcc30ca256eb00001b48e/53bbe9630b24d6f4ca256c3a000475b8!OpenDocument#Collection%2520District%2520(CD)).

Australian Capital Territory, Budget Statements. (2019). Education Directorate. Canberra; Author. [https://apps.treasury.act.gov.au/\\_data/assets/pdf\\_file/0007/1369789/F-Education-Directorate.pdf](https://apps.treasury.act.gov.au/_data/assets/pdf_file/0007/1369789/F-Education-Directorate.pdf).

Australian Council for Educational Research (ACER). (2018). *Scottish National Standardised Assessments: national report for academic year 2017 to 2018*. Edinburgh: Scottish Government. Accessed on 16 June 2020 on <https://www.gov.scot/publications/scottish-national-standardised-assessments-national-report-academic-year-2017-2018/pages/2/>.

ACER. (2020a). *Progressive achievement: tests, teaching resources and professional learning*. Accessed on 29 June 2020 on <https://www.acer.org/au/pat>.

ACER. (2020b). *Scottish National Standardised Assessments: national report for academic year 2018 to 2019*. Edinburgh: Scottish Government. Accessed on 16 June 2020 on <https://www.gov.scot/publications/scottish-national-standardised-assessments-national-report-academic-year-2018-2019/>

ACARA (2010). Narrative Marking Manual. Retrieved from: [https://www.nap.edu.au/resources/2010\\_Marking\\_Guide.pdf](https://www.nap.edu.au/resources/2010_Marking_Guide.pdf).

ACARA (2013a). *Persuasive Marking Manual*. Retrieved from: [https://www.nap.edu.au/docs/default-source/resources/2013\\_persuasive\\_writing\\_marking\\_guide.pdf](https://www.nap.edu.au/docs/default-source/resources/2013_persuasive_writing_marking_guide.pdf)

ACARA (2013b). *Submission to the Senate Inquiry, The Effectiveness of the National Assessment Program – Literacy and Numeracy*. [https://www.aph.gov.au/Parliamentary\\_Business/Committees/Senate/Education\\_and\\_Employment/Naplan13](https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/Naplan13).

ACARA (2017). The Australian National Assessment Program Literacy and Numeracy (NAPLAN) assessment framework: NAPLAN Online 2017-2018. <https://www.nap.edu.au/docs/default-source/default-document-library/naplan-assessment-framework.pdf?sfvrsn=2>

ACARA (2018a). About us. Retrieved from: <https://www.acara.edu.au/about-us>

ACARA (2018b). *Colmar Brunton Report*. Accessed on 1 July 2020 on <https://acaraweb.blob.core.windows.net/acaraweb/docs/default-source/assessment-and-reporting-publications/2018-naplan-online-parent-research.pdf?sfvrsn=2>.

ACARA (2018c). *International Comparative Study: The Australian Curriculum and The British Columbia New Curriculum*. Retrieved from: <https://www.australiancurriculum.edu.au/media/3923/ac-bcc-international-comparative-study-final.pdf>



ACARA (2018d) *International Comparative Study: The Australian Curriculum and The Singapore Curriculum*. Retrieved from: <https://www.australiancurriculum.edu.au/media/3924/ac-sc-international-comparative-study-final.pdf>

ACARA (2018e). *NAPLAN 2017 Technical report*. Sydney: ACARA. Accessed on 6 July 2020 on <https://www.nap.edu.au/results-and-reports/national-reports>.

ACARA (2018f). *NAPLAN Online Automated Scoring Research Program: Research Report*. Retrieved from: <https://www.nap.edu.au/docs/default-source/default-document-library/naplan-online-aes-research-report-final.pdf?sfvrsn=0>

ACARA. (2019a). *Australian Curriculum English. Level Description*. Retrieved from: <https://www.australiancurriculum.edu.au/f-10-curriculum/english/>

ACARA (2019b). *NAPLAN national report for 2019*. Sydney: Author. Accessed on 25 June 2020 on <https://nap.edu.au/docs/default-source/resources/naplan-2019-national-report.pdf?sfvrsn=2>.

ACARA (2020a). *About us*. Retrieved from: <https://www.acara.edu.au/about-us>

ACARA (2020b). *Assessment: NAPLAN*. Accessed on 27 June 2020 on <https://www.acara.edu.au/assessment/naplan>.

ACARA (2020c). *My School*. Retrieved from: <https://myschool.edu.au/>

ACARA (2020d). *NAPLAN*. <https://www.nap.edu.au/naplan>.

ACARA (2020e). *NAPLAN 2019 Technical report*. Sydney: ACARA. Accessed on 27 June 2020 on <https://www.nap.edu.au/results-and-reports/national-reports>.

ACARA (2020f). *NAPLAN – adjustments for students with disability*. Accessed on 27 June 2020 on <https://www.nap.edu.au/naplan/school-support/adjustments-for-students-with-disability>.

ACARA (2020g). *NAPLAN – Australian Curriculum*. Accessed on 29 June 2020 on <https://www.nap.edu.au/naplan/australian-curriculum>.

ACARA (2020h). *NAPLAN Online*. Accessed on 29 June 2020 on <https://www.nap.edu.au/online-assessment>.

ACARA (2020i). *NAPLAN – participation*. Accessed on 27 June 2020 on <https://www.nap.edu.au/information/faqs/naplan--participation>.

ACARA (2020j). *NAPLAN Reports*. Retrieved from: <https://reports.acara.edu.au/Home/Results>

ACARA (2020k). *Student reports*. Accessed on 27 July 2020 on <https://nap.edu.au/results-and-reports/student-reports>

ACARA (2020l). *Terms of Reference – Review of The Australian Curriculum F-10*. Retrieved from: [https://www.acara.edu.au/docs/default-source/curriculum/ac-review\\_terms-of-reference\\_website.pdf](https://www.acara.edu.au/docs/default-source/curriculum/ac-review_terms-of-reference_website.pdf)

- Angus, M., Olney, H. & Ainley, J. (2007). *In the balance: the future of Australia's primary schools*. Kaleen, ACT: Australian Primary Principals Association. Accessed on 6 July 2020 on <https://appa.asn.au/wp-content/uploads/2020/05/In-the-balance.pdf>
- Australia, Department of Education, Skills and Employment. (2010). *Media release: My School website launched*. <https://ministers.dese.gov.au/gillard/my-school-website-launched>.
- Australian Education Council. (1989). *The Hobart Declaration on Schooling*. Author. <http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Hobart-Declaration-on-Schooling-1989.aspx>.
- Australia, Parliament. (2008). Bills Digest 60, 2008-09 Australian Curriculum, Assessment and Reporting Authority Bill. [https://www.aph.gov.au/Parliamentary\\_Business/Bills\\_Legislation/bd/bd0809/09bd060](https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd0809/09bd060).
- Berge, K., Evensen, L.S., & Theygesen, R. (2016). The Wheel of Writing: a model of the writing domain for the teaching and assessing of writing as a key competency, *The Curriculum Journal*, 27:2, 172-189, DOI: 10.1080/09585176.2015.1129980
- Beuchert, L.V., & Nandrup, A.B. (2015). *The Danish National Tests – A Practical Guide*. Economic Working Papers 2014-2025. Retrieved from: <https://pdfs.semanticscholar.org/fc01/9a5636c68d3accd776820fe2638c599cac58.pdf>
- Bridgeman, B., & Ramineni, C. (2017). Design and evaluation of automated writing evaluation models: Relationships with writing in naturalistic settings. *Assessing Writing* (34): 62–71.
- Center on International Education Benchmarking (CIEB). (2020) *Japan: learning systems*. Accessed on 15 June 2020. <https://ncee.org/what-we-do/center-on-international-education-benchmarking/top-performing-countries/japan-overview/japan-instructional-systems/>.
- Collins, S. (2017). Government confirms primary schools to scrap National Standards. *New Zealand Herald*. Accessed on 23 June 2020 on [https://www.nzherald.co.nz/nz/news/article.cfm?c\\_id=1&objectid=11958067](https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11958067).
- Comber, B. (2012). Mandated literacy assessment and the reorganisation of teachers' work: Federal policy, local effects. *Critical Studies in Education*, 53(2), 119-136. <http://doi.org/10.1080/17508487.2012.672331>
- Council of Australian Governments. (2018). *National School Reform Agreement*. [https://docs.education.gov.au/system/files/doc/other/national\\_school\\_reform\\_agreement\\_8.pdf](https://docs.education.gov.au/system/files/doc/other/national_school_reform_agreement_8.pdf)
- Curriculum Corporation. (2005). *Statements of Learning for English*. Carlton South: Curriculum Corporation. Accessed on 29 June 2020 on [http://www.curriculum.edu.au/verve/\\_resources/SOL\\_English\\_Copyright\\_update2008\\_file.pdf](http://www.curriculum.edu.au/verve/_resources/SOL_English_Copyright_update2008_file.pdf)
- Curriculum Corporation. (2006). *Statements of Learning for Mathematics*. Carlton South: Curriculum Corporation. Accessed on 29 June 2020 on [http://www.curriculum.edu.au/verve/\\_resources/SOL\\_Mathematics\\_2006.pdf](http://www.curriculum.edu.au/verve/_resources/SOL_Mathematics_2006.pdf)
- Delandshere, G., & Petrosky, A.R. (1998). Assessment of Complex Performance: Limitations of Key Measurement Assumptions. *Educational Researcher*, 27(2), 14-24.

Dulfer, N., Polesel, J., & Rice, S. (2012). *The experience of education: The impacts of high stakes testing on school students and their families. An Educator's Perspective*. University of Western Sydney: Whitlam Institute.

Education Council. (2014). *Communiqué*. 31 October 2014. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiques%20and%20Media%20Releases/2014%20Communiques/Education%20Council%2031%20October%20Communique.pdf>

Education Council. (2015). *Communiqué*. Fifth Education Council meeting, 29 May 2015, Brisbane. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/EC%20Communiques%20and%20media%20releases/Education%20Council%2029%20May%202015%20-%20Communique.pdf>.

Education Council. (2016). *Media release: My School updated for 2016*. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Key%20Documents/Education%20Council%20media%20release%20-%20My%20School%20updated%20for%202016.pdf>.

Education Council. (2018). *Communiqué*. 13th April 2018. Adelaide. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiques%20and%20Media%20Releases/2018%20Media%20Releases/Education%20Council%20Communique%2013%20April%202018.pdf>

Education Council. (2019a). *Alice Springs (Mparntwe) Declaration*. Author. <http://www.educationcouncil.edu.au/Alice-Springs--Mparntwe--Education-Declaration.aspx>.

Education Council, (2019b). *Communiqué*. 28 June 2019, Melbourne. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiques%20and%20Media%20Releases/2019%20media%20releases/Education%20Council%20Communique%2028%20June%202019%20final.pdf>.

Education Council, (2019c). *Communiqué*. 12 December 2019, Alice Springs <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/EC%20Communiques%20and%20media%20releases/Education%20Council%20Communique%20-%2012%20December%202019.pdf>.

Education Quality and Accountability Office. (2007). *Framework Ontario Secondary School Literacy Test*. December Edition. Ontario, Canada.

Education Quality and Accountability Office. (2020). *EQAO's Technical Report for the 2017–2018 Assessments*. Toronto, Canada.

Education Services, NZ Ministry of Education. (2020). *PaCT information*. Accessed on 23 June 2020 on <https://services.education.govt.nz/schools/pact/information/>.

Electronic Assessment Tools for Teaching and Learning (e-asTTle). (2020). *E-asTTle basics*. Retrieved from: <https://e-asttle.tki.org.nz/About-e-asTTle/Basics>

Eliot, N., Ruggles Gere, A., Gibson, G., Toth, C., Whithaus, C & Presswood, A. (2013). *Uses and Limitations of Automated Writing Evaluation Software*. Council of Writing Program Administrators ComPILE Research Bibliographies, No. 23.

Finnish Education Evaluation Centre (FINEEC). (2020). *Learning outcomes evaluations*. Accessed on 24 June 2020 on <https://karvi.fi/en/>.

GL Assessment. (2020). *York Assessment of Reading for Comprehension*. Accessed on 29 June 2020 on <https://www.gl-assessment.co.uk/support/yarc-support/australia>.

Gorur, R. (2016). The performative politics of NAPLAN and My School. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (1st ed., 30-43). Oxon: Routledge.

Grønmo, L., Lindquist, M., Arora, A., & Mullis, I. (2013). *TIMSS 2015 Mathematics Framework*. [https://timssandpirls.bc.edu/timss2015/downloads/T15\\_FW\\_Chap1.pdf](https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap1.pdf)

Hardy, I. (2014). A logic of appropriation: enacting national testing (NAPLAN) in Australia. *Journal of Education Policy*, 29:1, 1-18, doi: [10.1080/02680939.2013.782425](https://doi.org/10.1080/02680939.2013.782425)

Harlen, W. (2005a). Teachers' summative practices and assessment for learning – tensions and synergies, *The Curriculum Journal*, 16 (2), 207-223, doi: [10.1080/09585170500136093](https://doi.org/10.1080/09585170500136093)

Harlen, W. (2005b) Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes, *Research Papers in Education*, 20 (3), 245-270, doi: [10.1080/02671520500193744](https://doi.org/10.1080/02671520500193744)

Hayward, E. (2018). Notes from a small country: teacher education, learning innovation and accountability Scotland. In Wyatt-Smith, C. & Adie, L. (eds.) *Innovation and accountability in teacher education*. Singapore: Springer, pp. 37-50. (doi:[10.1007/978-981-13-2026-2\\_3](https://doi.org/10.1007/978-981-13-2026-2_3))

Hong Kong Examinations and Assessment Authority. (2019). *Hong Kong National Report of Territory-wide System Assessment (TSA)*. Retrieved from: <http://www.bca.hkeaa.edu.hk/web/TSA/en/2019tsaReport/eng/TSA2019E.pdf>

Hong Kong Examinations and Assessment Authority. (2020). *Introduction*. Retrieved from: <http://www.bca.hkeaa.edu.hk/web/TSA/en/Introduction.html>

Houlberg, K., Andersen, V.N., Bjørnholt, B., Krassel, K.F & Pedersen, L.H. (2016). *Country Background Report- Denmark*. OECD Review of Policies to Improve the Effectiveness of Resource Use in Schools (Project no. 10932). Retrieved from [www.kora.dk](http://www.kora.dk)

Humphrey, S., & Heldsinger, S. (2019). Raters' perceptions of assessment criteria relevance. *Assessing Writing*, 41, 1-13.

Humphrey, S., & Heldsinger, S. (2014). *Common Structural Design Features of Rubrics May Represent a Threat to Validity*. *Educational Researcher*, 43(5), 253-263. DOI: 10.3102/0013189X14542154

Hutchinson, C. & Young, M. (2011). Assessment for learning in the accountability era: empirical evidence from Scotland, *Studies in Educational Evaluation*, 37, 62-70. <https://doi.org/10.1016/j.stueduc.2011.03.007>

Jeffery, J.V., Elf.N., Skarc, G.B & Campbell Wilcox, K. (2019). Writing development and education standards in cross-national perspective. *Writing and Pedagogy*, 10 (3), 333–370.

Kōichi, N. (2012). *Japan's drifting education system: the debate of Japan's academic decline*. Nippon.co. Accessed on 23 June 2020 on <https://www.nippon.com/en/in-depth/a00601/the-debate-over-japan's-academic-decline.html>.

Krogh, E. (2018). Crossing the Divide Between Writing Cultures. In Kristyan Spelman Miller & Marie Stevenson (eds.). *Transitions in writing* (pp.72-104). Leiden/Boston: Brill.

Kuramoto, N. & Koizumi, R. (2016). Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy & Practice*, 25, 415-433, DOI: 10.1080/0969594X.2016.1225667.

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29, 4-16.

Louden, W. (2019). NAPLAN reporting review: prepared for the COAG Education Council. Melbourne: Education Services Australia. Accessed on 1 July 2020 on <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Reports%20and%20publications/NAPLAN%20Reporting%20Review/Final%20Report.pdf>.

Lu, M., Turnbull, M., Wan, W.Y., Rickard, K., & Hamilton, L., (2017). *Are writing scores from online writing tests for primary students comparable to those from paper tests?* Centre for Education Statistics and Evaluation: Sydney, New South Wales.

Manitoba Education and Training. (2020a). *Assessment and Evaluation*. Retrieved from: <https://www.edu.gov.mb.ca/k12/assess/myreporting.html>

Manitoba Education and Training. (2020b). *Middle years assessment: Grade 8 English language arts: reading comprehension and expository writing: support document for teachers*. Winnipeg, Manitoba, Canada. ISBN: 978-0-7711-7516-9 (pdf)

Manitoba Education and Training. (2020c). *Provincial Results*. Retrieved from: <https://www.edu.gov.mb.ca/k12/assess/results/index.html>

Masters, G.N. & Forster, M. (1997a). *Literacy standards in Australia*. Melbourne: ACER. Accessed on 13 July 2020 on [https://research.acer.edu.au/cgi/viewcontent.cgi?article=1005&context=monitoring\\_learning](https://research.acer.edu.au/cgi/viewcontent.cgi?article=1005&context=monitoring_learning).

Masters, G.N. & Forster, M. (1997b). *Mapping literacy achievement: results of the 1996 National School English Literacy Survey*. Melbourne: ACER.

Matriculation Examination Board, Finland. (2020). *Matriculation Examination*. Accessed on 16 June 2020 on <https://www.ylioppilastutkinto.fi/en/matriculation-examination/the-examination>.

Matters, G. (2018). Queensland NAPLAN Review. Parent Perceptions Report. Accessed on 1 July 2020 on <https://qed.qld.gov.au/programsinitiatives/education/Documents/naplan-2018-parent-perceptions-report.pdf>.

McGaw, B., Loudon, W. & Wyatt-Smith. (2019). *NAPLAN review interim report*. Sydney: NAPLAN Review.

Messick, S. (1994). The interplay of evidence and consequences in validation of performance assessment. *Educational Researcher*, 23(2), 13-23.

Ministerial Council on Education, Employment, Training and Youth Affairs. (1999). *The Adelaide Declaration on National Goals for Schooling in the Twenty First Century*. Canberra: Author. <http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Adelaide-Declaration.aspx>.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2005). *Information statement*, 18th MCEETYA meeting, Canberra 12 May 2005 to 13 May 2005. [http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiqués%20and%20Media%20Releases/Previous%20Council%20info%20statements/MCEETYA%20meeting%20info%20statements/MC18\\_information\\_statement.pdf](http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiqués%20and%20Media%20Releases/Previous%20Council%20info%20statements/MCEETYA%20meeting%20info%20statements/MC18_information_statement.pdf).

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008a). *Melbourne Declaration on Educational Goals for Young Australians*. Canberra: Author. [http://www.curriculum.edu.au/verve/\\_resources/National\\_Declaration\\_on\\_the\\_Educational\\_Goals\\_for\\_Young\\_Australians.pdf](http://www.curriculum.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf).

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008b). *Media Release*. <https://ministers.dese.gov.au/gillard/ministerial-council-education-employment-training-and-youth-affairs>.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008c). *National Report on Schooling in Australia*. Author. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Reports%20and%20publications/Archive%20Publications/National%20Report/ANR%202008.pdf>.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2009a). *Communiqué*, 17th April 2009, Adelaide. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiqués%20and%20Media%20Releases/Previous%20Council%20info%20statements/MCEETYA%20meeting%20info%20statements/MC27%20Communique%20Inc%20ACER.pdf>.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2009b). *Principles and Protocols for Reporting on Schooling in Australia*. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Reports%20and%20publications/Publications/Measuring%20and%20reporting%20student%20performance/Principles%20and%20protocols%20for%20reporting%20on%20schooling%20in%20Australia.pdf>.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2011). *Communique*. Twelfth MCEECTYA Meeting, 8 July 2011, Melbourne. [http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiqués%20and%20Media%20Releases/Previous%20Council%20info%20statements/MCEECDYA%20meeting%20info%20statements/C12\\_Communique.pdf](http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiqués%20and%20Media%20Releases/Previous%20Council%20info%20statements/MCEECDYA%20meeting%20info%20statements/C12_Communique.pdf).

Ministry of Children and Education. (2020). *Examinations and Other Forms of Assessment*. Retrieved from: <https://eng.uvm.dk/primary-and-lower-secondary-education/the-folkeskole/examinations-and-other-forms-of-assessment>

Ministry of Education, Culture, Sports, Science and Technology – Japan (MEXT). 2020. *Principals guide Japan's educational system*. Accessed 15 June on <https://www.mext.go.jp/en/policy/education/overview/index.htm>.

Ministry of Education Singapore (2018). 'Learn for life' – preparing our students to excel beyond exam results. Accessed 27 June 2020 on <https://www.moe.gov.sg/news/press-releases/-learn-for-life---preparing-our-students-to-excel-beyond-exam-results>.

Ministry of Education Singapore. (2020a). *From primary to secondary education*. Accessed on 27 June 2020 on <https://www.moe.gov.sg/education/primary/from-primary-to-secondary-education>.

Ministry of Education Singapore. (2020b). *GCE 'A' Level curriculum*. Accessed on 15 June 2020 on <https://www.moe.gov.sg/education/pre-university/gce-a-level-curriculum>.

Ministry of Education Singapore. (2020c). *Primary school subjects and syllabuses*. Accessed on 15 June 2020 on <https://beta.moe.gov.sg/primary/curriculum/syllabus/>.

National Center for Education Statistics (NCES). (2017). *NAEP 2017 Writing Assessments*. Retrieved from: [https://nces.ed.gov/nationsreportcard/subject/writing/pdf/2017\\_writing\\_technical\\_summary.pdf](https://nces.ed.gov/nationsreportcard/subject/writing/pdf/2017_writing_technical_summary.pdf)

National Center for Education Statistics (NCES). (2019). *An Overview of NAEP*. Retrieved from [https://nces.ed.gov/nationsreportcard/subject/about/pdf/naep\\_overview\\_brochure\\_2018.pdf](https://nces.ed.gov/nationsreportcard/subject/about/pdf/naep_overview_brochure_2018.pdf)

National Center for Education Statistics (NCES). (2020a). *Assessment Schedule*. Retrieved from: <https://nces.ed.gov/nationsreportcard/about/calendar.aspx>

National Center for Education Statistics (NCES). (2020b). *Writing Interrater Agreement*. Retrieved from: [https://nces.ed.gov/nationsreportcard/tdw/scoring/scoring\\_within\\_wri.aspx](https://nces.ed.gov/nationsreportcard/tdw/scoring/scoring_within_wri.aspx)

National University of Singapore. (2020). *Singapore-Cambridge GCE 'A' Level: Admission requirements*. Accessed on 15 June 2020 on <http://www.nus.edu.sg/oam/apply-to-nus/singapore-cambridge-gce-a-level/admissions-requirements>.

New South Wales, Department of Education. (2018). *Strategic Plan 2018-2022*. Sydney: Author. <https://education.nsw.gov.au/about-us/strategies-and-reports/strategic-plan#Our5>.

Nelson. (2020). *PM Benchmark*. Accessed on 29 June 2020 on <https://cengage.com.au/primary/browse-series/pm/pm-benchmark>.

New York State Education Department. (2020). *Educator Guide to the 2020 Grades 3–8 English Language Arts Tests*. Retrieved from: <https://www.engageny.org>

NZCER, (2020a). *Assessment Resource Banks*. Accessed on 26 June 2020 on <https://arbs.nzcer.org.nz>.

NZCER. (2020b). *Progressive Achievement Tests (PATs)*. Accessed on 23 June 2020 on <https://www.nzcer.org.nz/tests/pats>.

NZ Ministry of Education. (2020a). *Assessment online*. Accessed on 15 June 2020 on <http://assessment.tki.org.nz>.

NZ Ministry of Education. (2020b). *e-asTTle*. Accessed on 23 June 2020 on <https://e-asttle.tki.org.nz>.

NZ Ministry of Education. (2020c). *The national administration guidelines*. Accessed on 26 June 2020 on <https://www.education.govt.nz/our-work/legislation/nags/>.

Oates, T. (2015). *Finnish fairy stories*. Cambridge: Cambridge Assessment.

Official Norwegian Report (green paper) (2002). *First class from the first class – Proposal for a framework for a national quality assessment system of Norwegian basic education*. Retrieved from <https://www.regjeringen.no/no/dokumenter/nou-2002-10/id145378/sec5?q=grunnleggende#KAP3-4-1-P3>

OECD. (2001). *Knowledge and skills for life: first results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

OECD. (2004). *Learning for tomorrow's world: first results from PISA 2003*. Paris: OECD.

OECD. (2007). *PISA 2006: science competencies for tomorrow's world, Volume 1: analysis*. Paris: OECD.

OECD. (2010). *PISA 2009 results: What students know and can do – student performance in reading, mathematics and science, Volume I*. Paris: OECD.

OECD (2019a), *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/b25efab8-en>.

OECD. (2019b). *PISA 2018 results: what students know and can do, Volume I*, Paris: Author.

OECD (2020), *Reading performance (PISA) (indicator)*. <https://doi.org/10.1787/79913c69-en> (Accessed on 25 June 2020) <https://data.oecd.org/pisa/reading-performance-pisa.htm>

Office of Qualifications and Examinations (OFQUAL). (2020). Summary of changes to GCSEs from 2015. Accessed on 21 July 2020 on <https://www.gov.uk/government/organisations/ofqual>.

Okabe, T., Tose, N., & Nishimura, K. (1999). *Bunsuu ga dekinai daigakusei* [University students who cannot perform calculations using fractions]. Tokyo: Toyo Keizai.

Ontario Education Quality and Accountability Office. (2020a) *Everything you need to know about EQAQ elementary assessments*. Accessed 15 June on <https://www.eqao.com/en/assessments/communication-docs/guide-elementary-assessments-english.pdf>.

Ontario Education Quality and Accountability Office. (2020b). *Upcoming changes to EQAO's assessments and reports*. Accessed 15 June on [https://www.eqao.com/en/about\\_eqao/modernization/Pages/memo-changes-eqao-assessments-reports-2019.aspx](https://www.eqao.com/en/about_eqao/modernization/Pages/memo-changes-eqao-assessments-reports-2019.aspx).

Parkin, C. & Parkin, C. (2011). *PROBE2 Reading Comprehension Assessment*. Upper Hutt, NZ: Triune Initiatives. Accessed on 29 June 2020 on <https://comprehenz.com/resources-all-resources/resources-assessment/probe-2-reading-comprehension-assessment/>.

Perelman, L. (2018). *Towards A New NAPLAN: Testing to the Teaching*. Surry Hills, Sydney. ISBN 978-0-3482555-2-9.

Queensland Curriculum and Assessment Authority. (2019). *Retrospective: 2018 Queensland Core Skills Test Writing Task*. Accessed on 19 July on [https://www.qcaa.qld.edu.au/downloads/senior/qcs\\_retro\\_18\\_3.pdf](https://www.qcaa.qld.edu.au/downloads/senior/qcs_retro_18_3.pdf).

Queensland, Department of Education. (2019). *Service Delivery Statements*. Brisbane: Author. <https://budget.qld.gov.au/files/2019-20%20DoE%20SDS.pdf>.



Reedy, D. (undated). Independent review of the Scottish National Standardised Assessments at Primary 1. Accessed on 16 June 2020 on <https://www.gov.scot/binaries/content/documents/govscot/publications/progress-report/2019/06/scottish-national-standardised-assessments-review-2019/documents/independent-review-of-the-scottish-national-standardised-assessments-at-primary-1/independent-review-of-the-scottish-national-standardised-assessments-at-primary-1/govscot%3Adocument/Independent%2BReview%2Bof%2Bthe%2BScottish%2BNational%2BStandardised%2BAssessments%2Bat%2BPrimary%2B1.pdf>

Rezaei, A.R & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15 (1), 18-39.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2): 119–44.

Sadler, D.R. (2009). Indeterminacy in the use of preset criteria for assessment and grading, *Assessment & Evaluation in Higher Education*, 34:2, 159-179, DOI: 10.1080/02602930801956059

Sahlberg, P. (2015). *Finnish lessons 2.0: what can the world learn from educational change in Finland?* New York: Teachers College Press.

Sahlberg, P. (2016). The global education reform movement and its impact on schooling. In (Eds.) Mundy, K., Green, A., Lingard, B. & Verger, A. *The handbook of global education policy*. doi:10.1002/9781118468005.ch7. Chichester: John Wiley & Sons, pp. 145-161.

SEAMEO INNOTECH, (2015). Assessment Systems in Southeast Asia: Models, Successes and Challenges. Retrieved from: <http://www.seameo-innotech.org>

Singapore Examinations and Assessment Board (2020). *Your trusted authority in examinations and assessment*. Accessed on 15 June 2020 on <https://www.seab.gov.sg/home/#>.

Scottish Government. (2016). National improvement framework for Scottish education: 2016 evidence report. Edinburgh: Author. Accessed on 16 June 2020 on <https://www.gov.scot/publications/national-improvement-framework-scottish-education-2016-evidence-report/>.

Shermis, M. (2014). The challenges of emulating human behavior in writing assessment. *Assessing Writing*, 22, 91-99.

Shewbridge, C., Jang, E., Matthews, P., & Santiago, P. (2011). *OECD Reviews of Evaluation and Assessment in Education Denmark Main Conclusions*. Retrieved from: [www.oecd.org/edu/evaluationpolicy](http://www.oecd.org/edu/evaluationpolicy)

Singapore Examinations and Assessment Board (SEAB). (2020). *Behind the Scene: Key Exam Processes*. Retrieved from: <https://www.seab.gov.sg/home/examinations/psle/behind-the-scene>

Skar, G. (2017). *The Norwegian National Sample-Based Writing Test 2016: Technical Report*. Retrieved from: <http://www.skrivesenteret.no/uploads/files/Skriveproven2017/NSBWT2017.pdf>

Skar, G. B. & Jølle, L. J. (2017). Teachers as raters: An investigation of a long-term writing assessment program. *L1-Educational Studies in Language and Literature*, 17, 1-30. <https://doi.org/10.17239/L1ESLL-2017.17.01.06>

Thompson, G. (2013). NAPLAN, My School and accountability: Teacher perceptions of the effects of testing. *The International Education Journal: Comparative Perspectives*, 12(2), 62-84. Retrieved from: [www.iejcomparative.org](http://www.iejcomparative.org)

Thomson, S., Hillman, K., Schmid, M., Rodrigues, S., and Fullarton, J. (2017a). *Highlights from PIRLS 2016 Australia's perspective*. Melbourne: Australian Council for Educational Research. <https://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=pirls>

Thomson, S., Wernert, N., O'Grady, E., & Rodrigues, S. (2017b). *TIMSS 2015: Reporting Australia's results*. Melbourne: Australian Council for Educational Research. [https://research.acer.edu.au/cgi/viewcontent.cgi?article=1002&context=timss\\_2015](https://research.acer.edu.au/cgi/viewcontent.cgi?article=1002&context=timss_2015)

Thomson, S., De Bortoli, L., Underwood, C., & Schmid, M. (2019) *PISA 2018: Reporting Australia's Results. Volume I Student Performance*. Melbourne: Australian Council for Educational Research. <https://research.acer.edu.au/cgi/viewcontent.cgi?article=1035&context=ozpisa>

Today Online. (2020). *Secondary school streaming to be abolished* [in Singapore]. Accessed on 15 June 2020 from <https://www.todayonline.com/singapore/secondary-school-streaming-be-abolished-2024-replaced-subject-based-banding>.

Track One Studio. (2020). *Learning analytics suite*. Accessed on 29 June 2020 on <https://www.trackonestudio.com>.

UK Department of Education. (2017). *Primary assessment in England: equalities impact assessment*. London: Author. Accessed on 22 June 2020 on [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/644717/Primary\\_assessment\\_in\\_England\\_-\\_EIA.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/644717/Primary_assessment_in_England_-_EIA.pdf).

UNICEF & SEAMEO. (2019a). *SEA-PLM 2019 Assessment Framework (1st ed.)*. Bangkok, Thailand: United Nations Children's Fund (UNICEF) & Southeast Asian Ministers of Education Organization (SEAMEO) – SEA-PLM Secretariat.

UNICEF & SEAMEO. (2019b). *SEA-PLM 2019 Technical Standards*. Bangkok, Thailand: United Nations Children's Fund (UNICEF) & Southeast Asian Ministers of Education Organization (SEAMEO) – SEA-PLM Secretariat.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2019). *The promise of large-scale learning assessments Acknowledging limits to unlock opportunities*. Paris, France. ISBN 978-92-3-100333-2

University of Otago. (2020). *National Monitoring Study of Student Achievement*. Accessed on 23 June 2020 on <https://nmssa.otago.ac.nz/index.htm>.

University of Western Australia. (2020). *BASE Australia*. Accessed on 29 June 2020 on <http://www.education.uwa.edu.au/base>.

UNSW Global. (2020a). *Educational assessments: how Reach and ICAS are different*. Accessed on 29 June 2020 on <https://www.unswglobal.unsw.edu.au/educational-assessments/campaigns/reach-and-icas/>.

UNSW Global. (2020b). *Educational assessments: ICAS*. Accessed on 29 June 2020 on <https://www.unswglobal.unsw.edu.au/educational-assessments/products/icas-assessments/>

Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms, *Educational Review*, 71(1), 5-30, doi: 10.1080/00131911.2019.1522045

Victoria, Department of Education and Training. (2017). *Differentiated School Performance Method 2019*. Melbourne: Author. [https://www.education.vic.gov.au/Documents/school/teachers/management/improvement/2019\\_dspm\\_measuresguide.pdf](https://www.education.vic.gov.au/Documents/school/teachers/management/improvement/2019_dspm_measuresguide.pdf).

Victoria, Department of Education and Training. (2019). 2018-19 Report on Operations. Melbourne: Author. <https://www.education.vic.gov.au/Documents/about/department/2018-19-report-of-operations.pdf>.

Victoria, Department of Education and Training. (2019). *Fact sheet: Differentiated support for school improvement*. Melbourne: Author. <https://www.education.vic.gov.au/Documents/about/educationstate/differsupportedstatefactsheet.pdf>.

Wikipedia. (2020). *Ecological fallacy*. Accessed on 27 June 2020 on [https://en.wikipedia.org/wiki/Ecological\\_fallacy](https://en.wikipedia.org/wiki/Ecological_fallacy).

Wyatt-Smith, C. & Adie, L. (2020 forthcoming). The development of students' evaluative expertise: Enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*. [doi.org/10.1080/00220272.2019.1624831](https://doi.org/10.1080/00220272.2019.1624831)

