



UNSW
Gonski Institute
for Education

PUTTING STUDENTS FIRST: Moving on from NAPLAN to a new educational assessment system



Rachel Wilson
Adrian Piccoli
with
Andy Hargreaves
Pak Tee Ng
Pasi Sahlberg



UNSW
SYDNEY

Authors

Rachel Wilson is Associate Professor in educational assessment, evaluation and research methods, at the University of Sydney. Rachel's background is in psychology, teaching and research methodology. She is a researcher at the university's Centre for Educational Measurement (CEMA). She has broad interests across educational policy and practice and is interested in system level reform and innovation. In this work she strives to maintain focus on students and their broader development, she is author of *Understanding Emotional Development* (2016).

Adrian Piccoli is the former Director of the UNSW Gonski Institute for Education and previously served as a Member of NSW Parliament for 19 years and as the NSW Minister for Education from 2011 until 2017. He is a Fellow of the Australian Council for Educational Leaders and Honorary Professor at the UNSW. Adrian is author of *12 Ways Your Child Can Get the Best Out of School* (2019).

Andy Hargreaves is Director of CHENINE (Change, Engagement and Innovation in Education at the University of Ottawa), Research Professor in the Lynch School of Education and Human Development at Boston College and holds Visiting or Honorary Professorships at Hong Kong University and Swansea University. He is Past President of the International Congress of School Effectiveness and Improvement, Commissioning Editor for the Journal of Professional Capital and Community, recent Adviser in Education to the Premier of Ontario and currently to the First Minister of Scotland. He is co-founder and President of the ARC Education Project: a group of nations committed to broadly defined excellence, equity, wellbeing, inclusion, democracy and human rights in education.

Pak Tee Ng is a Singaporean educator at the National Institute of Education (NIE), Nanyang Technological University. He is passionate about the development of schools and teacher leaders. His main work is in educational change, policy and leadership. Pak Tee has spoken at many global events such as Google Global Education Symposium, International Baccalaureate Global Conference, and Scottish Learning Festival. Also, he is currently a member of Scotland's International Council of Education Advisers. His greatest reward is a fraternity of students who are educational leaders themselves and who pay it forward through their educational leadership.

Pasi Sahlberg is Professor of Education Policy at UNSW Sydney. He previously served as a Director General at Finland's Ministry of Education and as a Lead Education Specialist at the World Bank and the European Commission. He is author of *Finnish Lessons 3.0: What can the world learn from educational change in Finland?* (2021), and *In Teachers We Trust: The Finnish way to world-class schools* (2021, with Tim Walker).

Gonski Institute for Education

UNSW Sydney NSW 2052 Australia

E gonski@unsw.edu.au

W gie.unsw.edu.au

© UNSW Sydney 2021

Cover photo by [Nosnibor137/Bigstock.com](https://www.bigstock.com/author/nosnibor137/)

Suggested citation:

Wilson, R., Piccoli, A., Hargreaves, A., Ng, P. T., & Sahlberg, P. (2021). *Putting Students First: Moving on from NAPLAN to a new educational assessment system (The Gonski Institute Policy Paper #2-2021)*. Sydney: UNSW Gonski Institute.

PART 1: PUTTING STUDENTS FIRST	1
1 Introduction	1
2 Why do we need a new assessment system?	2
3 A vision for Australia	4
4 From census-based to sample-based student assessment	7
5 From measuring to learning	9
6 From accountability to trust-based responsibility	10
7 A new national assessment system	10
PART 2: PUTTING NAPLAN TO THE TEST	13
1 NAPLAN: Contexts and background	13
2 How does NAPLAN meet the assessment challenges?	19
Challenge 1: Aligning educational standards and student assessment	22
Challenge 2: Balance between summative and formative assessments	25
Challenge 3: Balance between external testing and teacher-led assessments	28
Challenge 4: Fair assessment of all students	29
Challenge 5: Large-scale assessments that are instructionally useful	32
Challenge 6: Fair assessment and marking across schools	35
Challenge 7: Informative reporting and proper use of assessment	37
Challenge 8: Ensuring assessment is informed, valued and of optimal utility to teachers	38
PART 3: INTERNATIONAL PERSPECTIVES	41
1 Scotland	41
2 Singapore	47
3 Ontario	51
4 Finland	57
PART 4: CONCLUSIONS	63
REFERENCES	67
ANNEXES	75

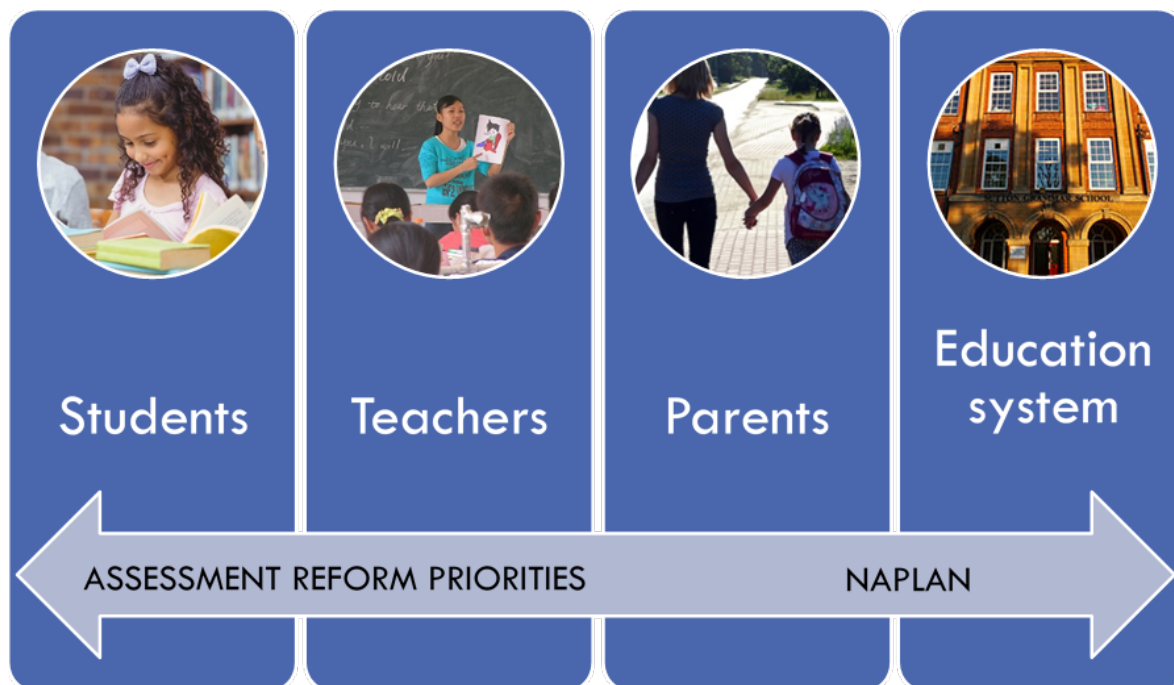
PART 1: PUTTING STUDENTS FIRST

1 Introduction

This report follows the Gonski Institutes [submission](#) to the Education Council of Australia - Government (COAG) review of the current approach to the presentation of the National Assessment Program - Literacy and Numeracy (NAPLAN) in March 2019 which provided an analysis of NAPLAN with particular consideration of how it is experienced by students, teachers and parents - and evaluation of it against a set of contemporary challenges to national assessment systems worldwide.

In this present report we start by outlining a new set of aims for Australia's national educational assessment system, that is oriented to student, teacher, parent and system needs and utilises contemporary understanding of assessment. These aims are summarised in Figure 1. This contrasts with the education reform perspectives of the early 2000's, when NAPLAN was conceived, which prioritised system monitoring and accountability at the cost of other aims and outcomes. The needs of all stakeholders, not just those of administration, are central to the new aims. They are in priority order, with the student at the apex.

Figure 1. Changing national assessment priorities



With clear new aims in mind, we conceptualise a new National Assessment System (NAS) that addresses the needs of students, teachers, parents, schools and educational systems. In meeting the needs of these primary and secondary stakeholders a new NAS must also address the challenges that face all modern educational assessment systems. These were cited in the Gonski Institute’s submission to the COAG mentioned above. Drawing on those findings, and with new aims in mind, this report goes on to provide: (1) the argument for new aims and reform; (2) an analysis of the needs of stakeholders; (3) an outline the key features of a new system; and (4) clarity on how this will shift and differ from the current model. Part 1 of this report concludes with a series of recommendations for progress toward a new National Assessment System.

2 Why do we need a new assessment system?

We start by considering recent perspectives on NAPLAN and clarifying their orientation to particular aims and purposes. We then put forward a new perspective which reorients Australia’s national assessment system by considering how it might better serve students, teachers, parents and the broader education system in that order of priority.

The recent [review of NAPLAN](#) (The McGaw Review) commissioned by the NSW, Victorian, Queensland and ACT governments, outlined NAPLAN’s “five purposes for national standardised testing ... endorsed by the ministerial council” as:

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

(McGaw et al., 2020)

While many would contest that there was ever any clarity on NAPLAN’s aims and purposes, these five points do reflect the orientation NAPLAN took and the language employed conveys the functional and instrumental approach to its design.

Although *the McGaw Review* outlined these aims in the [NAPLAN Review Final Report](#), there was no evaluative judgment of how NAPLAN achieved against them. Rather, the review produced a wide-ranging discussion of the aims and uses of assessment, national and international practices and only two chapters examining the “quality” of the NAPLAN digital and writing tests. Thus, the review failed to evaluate NAPLAN directly against its stated aims. Using available research evidence on NAPLAN, in Table 1 we provide an account of what such an evaluation might have looked like. Evidence supporting the conclusions in Table 1 is provided throughout this report.

Table 1. Evaluation of NAPLAN against five aims

Monitoring national, state and territory programs and policies	System accountability and performance	School improvement	Individual student learning and growth	Information for parents on school and student performance
<ul style="list-style-type: none"> • National literacy and numeracy data collection has been achieved • Monitoring is not aligned with some national goals, and may be hinder development of “excellence and equity” and “confident and creative” students • Potential for monitoring to address school funding and resourcing inequities is unrealised • Monitoring has not yet been linked to workforce data • Monitoring has not redressed educational inequity • Potential for long-term analysis and planning from monitoring has not been realised 	<ul style="list-style-type: none"> • Inequity among states and territories has not reduced, despite NAPLANs role in commonwealth accountabilities frameworks • Early linking commonwealth to state funding to NAPLAN was unsuccessful • Performance has improved significantly but not substantively in some tests, there are some evident declines in writing • Linking NAPLAN to <i>MySchool</i> produced a range of consequences, including negative ones • Potential to lift performance via formative AfL has been missed 	<ul style="list-style-type: none"> • NAPLAN is notably unsuitable, and not designed, for formative/AfL purposes • Timing of NAPLAN does not allow for an effective school improvement cycle • Inconsistent and contradictory messages have stymied the potential for school improvement to be driven by NAPLAN • The narrowing of the taught curriculum, in response to NAPLAN, hinders holistic school improvement 	<ul style="list-style-type: none"> • NAPLAN participation has declined over the decade - so for increasing numbers this aim is not addressed • NAPLAN tests are not suitable for many students, including many Indigenous students and students with disabilities • Although NAPLAN is designed to identify students not meeting minimum national benchmarks, trends show no improvement for those disadvantaged students • The timing and design of NAPLAN misses opportunities to strengthen student learning achievement and growth 	<ul style="list-style-type: none"> • Some parents are satisfied with NAPLAN reports • Parents are concerned about several issues related to test fairness; including: disadvantage for less resourced schools, stress on students, and the exclusion of some students from the test • Parents report concerns as to whether NAPLAN accurately reflects all students’ performance • Parents perceive the usefulness of NAPLAN as limited because of the long time period between sitting the tests and receiving the results

In sum, Table 1 shows that even when NAPLAN is assessed against these dated and poorly oriented aims, it fails to live up to expectations. Whilst there are a couple of bright spots, the introduction of a national assessment system and the satisfaction of some parents, there are several areas of concern in *the McGaw Review* – including those related to unintended and negative outcomes. There are also many missed opportunities to develop an assessment system that lives up to the aims of school improvement and student learning.

Had *the McGaw Review* evaluated NAPLAN against its five aims and made similar conclusions based on the evidence, that would have provided an evaluation on whether NAPLAN was doing what it was supposed to do when it started over a decade ago. We argue that those aims are now outdated and inappropriate, they take a narrow perspective on NAPLAN that is constrained to the type of thinking that dominated educational discourse during the 1990’s and early 2000s. This perspective led to the creation of NAPLAN and the *MySchool* Website; and has since proven unproductive.

3 A vision for Australia

In this review, we take a different and much broader approach to reviewing NAPLAN. Rather than judge whether NAPLAN is achieving its originally stated goals, we consider what a new Australian National Assessment System should look like. To do this we evaluated NAPLAN against nine challenges for any national assessment system (OECD, 2013). These reflect a more contemporary perspective on what modern educational assessment policy and systems can achieve.

Those objectives are:

1. Aligning educational standards and student assessment
2. Finding a balance between summative and formative assessment
3. Balancing external assessments and teacher-based assessments in the assessment of learning
4. Developing fair assessments to all student groups
5. Designing large-scale assessments that are instructionally useful
6. Ensuring fairness in assessment and marking across schools
7. Securing informative reporting of student assessment results
8. Maximising the potential for student learning through the assessment process
9. Ensuring the assessment is informed, valued and of optimal utility to the teaching profession

Using abovementioned approach, Part 2 of this document provides substantial evidence that NAPLAN is, at worst, deeply problematic, imposing a range of undesirable dynamics and consequences. At best, it represents a lost opportunity to harness the potential of assessment to strengthen the fabric of education in schools, provide a valuable resource for teacher professionalism, drive student learning outcomes and meet the national education goals for young Australians stated in the Alice Springs (Mparntwe) Education Declaration.

We also looked at how some of the leading education systems have organised student assessment today. Evidence from case studies of Finland, Ontario, Scotland and Singapore indicate that they are heading in the opposite direction to which Australia has found itself. They have all moved away from high stakes standardised testing and using test-based accountability to drive improvement. Instead, they are focusing energy, resources and assessment budgets to focus on Assessment for Learning (AfL) and building teacher professional capacity to harness the potential of assessment to drive student learning and school improvement. Importantly, all of those jurisdictions have recognised and taken action to redress the negative education and wellbeing consequences to some students of too much pressure to compete in standardised tests, particularly at young ages.

Federal Minister for Education, Alan Tudge, recently announced: “We must protect NAPLAN, and not give in to those who call for less accountability and less information for teachers and parents.” (APH, March 2021). He also acknowledged the need to “embed formative assessment in every classroom in the country” and “develop a repository of proven assessment tools that teachers and parents can use, in classrooms and at home”. Unfortunately, reform based on

tweaking existing NAPLAN, oriented to out-dated educational aims and purposes, will mean that Australia continues to spend on a system that is based on the wrong drivers. Adding to the current system, ignoring the negative side effects it has on many students, families, schools, and the entire system, and insisting it still holds potential to drive learning, denies the history and research evidence on NAPLAN. Tinkering with the past system is not enough. Rather, a substantial re-think is required.

This report offers a new vision that puts the primary focus back on what students, their teachers, parents and schools need from a national assessment system. Ultimately education systems are designed to serve students, yet student needs and experiences are not often part of ‘the logics’ of educational systems and their architecture. To consider what an optimal national assessment system should look like when we place students at the centre, we ask four questions:

1. What do students need from a national assessment system?
2. What information do teachers and schools need to support students?
3. What information do parents need to support their children and schools?
4. What is the necessary minimum information that governments need for accountability purposes and to support all of the above?

Table 2. Student-focused drivers for a national assessment system (NAS)

What do students need?	What do teachers and schools need to support students?	What do parents need to support their children and schools?	What do the systems need for monitoring and public accountability?
Developmentally appropriate low-stakes assessments, format and timing that does no harm and takes into account individual differences.	Assessments that are formative, flexible, timely and complement teachers’ own assessments.	Clear, written and visual reporting on their child’s learning throughout the school year which strengthens parents’ abilities to support their child’s teachers and their learning at home.	Develop and maintain an in-school, on-demand, formative assessment resource system, developed in partnership with the teaching profession that promotes student learning.
Assessment of what students have learned through a broad and integrated curriculum.	Assessments that inform the ongoing teaching and learning for both students and teachers.	Reporting on what a child can and cannot do across a broad curriculum relative to a series of validated education standards and benchmarks.	Undertake sample-based national assessments in Years 4, 6 and 8 and 10 that are summative for system monitoring, system-level accountability and system level decision making.
Assessment timing and mode that are integrated into teaching and learning, and that promote creativity and wellbeing.	Quality assured student assessment resources that help teachers to evaluate how their students learn and improve teaching based on that feedback.	Confidence that the assessment system will do no harm to their child, and that it will maximise the learning and wellbeing outcomes and holistic development of their child.	Undertake transparent system-level analysis of sample-based assessment data, that enables more comprehensive policy development including educational equity.
A national assessment system that accurately guides teachers, schools and systems to make decisions in support of students learning and wellbeing.	A national assessment system that is informed by the teaching profession, that attracts professional trust, involvement and respect and is useful to teaching and learning.	A national assessment system that raises public confidence in education by information about the education system performance and progress towards the set national goals.	A national assessment system that collects and utilises minimum required assessment data and analysis for public accountability and quality monitoring purposes.

Addressing these questions provides a student-focused national assessment system that exists in many high-performing education systems. Table 2 answers these questions and provides an analysis of the requirements for each of these stakeholder groups, laying out foundational principles for a productive and cost-effective assessment system that is oriented to student, teacher, parent and system needs.

Based on the stakeholder needs we recommend that a hybrid, three-part National Assessment System to be developed to address three key purposes:

- a) **Teaching and learning purposes:** The Australian national assessment body should develop a validated, formative classroom-based Assessment Resource System (ARS), including a library of curated, quality assessment tasks and a national item bank. The ARS will be used to provide formative, teacher-led assessment for years 3 through to year 10, carefully linked to the curriculum, informed by the teaching profession and integrated with teaching and learning. These resources are used by teachers at the time of their choosing. The ARS should support teachers, complement other teacher-led assessment and enrich current student reports. All the resources are mapped to and closely aligned with the Australian Curriculum and benchmarked to national standards.
- b) **Parent reporting purposes:** Results from formative classroom-based assessments using the ARS can be included on student reports and made available to parents in a systematic, visual and understandable format that reports on students' performance in various areas of knowledge and skills over time.
- c) **System monitoring purposes:** Annual national sample-based testing of students should be conducted for years 4, 6, 8 and 10. Curriculum areas to be included in these assessments should alternate from year to year according to 4-year national assessment plans. Data from summative assessments will be used to create an education system report card, without references to individual schools. The summative sample-based year 4, 6, 8 and 10 assessment system will also collect systematic data of students and their backgrounds to better understanding how different factors affect student learning. Annual reporting will include trends over time and disaggregation of results by key equity groups, in order to address national educational goals. This will create much richer and more comprehensive reporting on the Australian education system.

Figure 2. Contrasting current NAPLAN and potential reformed national assessment system

NAPLAN								
	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Formative Assessment for Learning	Currently non-systematic, inadequately unsupported, and often marginalised by high-stakes NAPLAN							
National summative sample-based testing				Science, civics & citizenship and ICT				Science, civics & citizenship and ICT
National summative census-based testing	Literacy Numeracy		Literacy Numeracy		Literacy Numeracy		Literacy Numeracy	
New National Assessment System (NAS)								
	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Formative Assessment for Learning tool available for all schools	Supported by an Assessment Resource System, broad curriculum coverage, values and helps teachers, closely aligned to AC and national standards, enriches student reports – optimises teaching and learning							
National summative sample-based testing		Literacy Numeracy + other domains		Literacy Numeracy + other domains		Literacy Numeracy + other domains		Literacy Numeracy + other domains

Reorientation toward the needs of students and refocussing on the potential of assessment to drive learning, results in several key differences between the old and new system. These include: a shift from standardised census-based testing to sample-based testing, that is primarily due to refocusing assessment for learning rather than assessment for summative measurement. There are also shifts from a testing focus to a broader assessment system which can scaffold, support and provide resources for teachers. Finally, there is a notable shift from NAPLAN’s accountability focus, toward a system which develops trust and invests in capacity building.

4 From census-based to sample-based student assessment

The central principle in this redesigned student assessment model is a system that combines sample-based assessments with a comprehensive program of classroom-based and teacher-led assessments. The McGaw Review argued that sample-based testing in Australia is not an option. However, the review failed to recognise the significance of modern classroom assessment practices for deeper learning – and the potential to align them to national standards. The review erroneously suggested that census-based testing alone is capable of driving school improvement and individual student learning achievement and growth; both of which these are noted weaknesses with NAPLAN’s census-based model. In fact, research clearly suggests that it is teacher-led assessment in the classroom that holds the greatest potential for meeting these aims, because such assessment can be fully embedded into teaching and learning on one hand and support school improvement on the other hand.

A full analysis of the strengths and weaknesses of sample- and census-based assessments is not complete without consideration of teacher-led classroom assessment practices, which hold the strongest potential for driving student learning. Comparing census- and sample-based testing without consideration of a properly developed system supporting classroom assessment, is misleading. Internationally, national assessment policies are increasingly focussing on developing systems and supports for this powerful type of student assessment.

Recent developments with assessment libraries and test-item banks mean that these classroom assessments, while not completed under controlled test conditions, can nevertheless provide important information of how children are learning relative to their peers, current curriculum and national standards or benchmarks. Inclusion of 'classroom assessment' in Table 3 provides a more balanced perspective and makes it clear that a combination of sample-based and teacher-led classroom assessments provide full, and many would argue improved, coverage of the purposes of standardised assessment.

Table 3. Purposes of different student assessment models assessments (expanded from the McGaw et al. [2020])

Purpose of national standardised assessment	Census-based	Sample-based	Classroom-based
Monitoring purposes towards national goals			
• National, jurisdictional and system estimates of student 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 performance	✓	✗	✓
• Relative school performance	✓	✗	✓

Not only does sample-based testing, in combination with strengthened classroom assessment, meet the wider range of assessment purposes; it provides a preferable model to current NAPLAN census-based testing because:

- Shifting to a sample-based assessment method removes the competitive aspects of testing. High-stakes effects, due to census-based testing and publishing results on *MySchool* website, are removed and many of the unintended negative consequences of NAPLAN are alleviated.
- Cost savings can be invested into strengthening classroom assessment.

5 From measuring to learning

Educational assessment experts suggest that we need to rethink the relationship between student assessment and learning. For example, Baird and colleagues conclude that *“If assessments are to serve the goals of education, then theories of learning and assessment should be developing more closely with each other”* (Baird et al., 2017). Regrettably, NAPLAN was developed as a summative measure of learning and its design was not oriented to theories of learning. A new assessment system must take up the challenge and be designed better for student learning purposes.

Developments in this field have been rapid over the last decade, and they focus primarily on how assessment can be embedded within the learning process. This means that a new system must promote high quality classroom assessment. Our proposal includes supporting this through:

- Engagement and consultation with teachers in the development of the system
- Development of a national Assessment Resource System (ARS)
- A program of support and professional learning for the effective implementation of the national system
- Continued opportunities for teachers to contribute to ARS and professional learning initiatives, and for the profession to be consulted for ongoing review and evaluation of the national assessment system.

Figure 3. Proposed new national assessment system



6 From accountability to trust-based responsibility

NAPLAN has a strong focus on accountability, both at the school and system level. Prioritising accountability resulted in neglect of other aspects and detracted from opportunities to focus on learning. A new national assessment system can rectify that. By designing a system focused on learning, a new NAS can live up to the promises that NAPLAN failed to deliver on. To focus on learning effectively we must think of the student's perspective.

Any fully formed national assessment system should include frameworks, resourcing, and support for teacher lead classroom-based assessment, so that assessment can drive learning and become an integrated and comfortable part of students' efforts across the full school year. The NAPLAN system, with its focus on tests in a single week, has little hope of matching the power of whole school year-round classroom assessment.

Furthermore, by moving away from the high-stakes test-based accountability model we are able to avoid many of the negative consequences that are associated with NAPLAN. By shifting from one-time standardised testing to continuous classroom-based assessment, there is the potential to reap many benefits. One of those is the strengthening of student and teacher relationships through the assessment process. Assessment that is led by teachers enables teachers to understand their students better and to build that knowledge into future teaching and learning.

School assessment plays a role in educational culture. Having an assessment system that barely considers teachers, never mind utilising the potential to drive learning, presents as a poor model for education. This neglect of teachers within the current NAPLAN system helps explain its failings. The national framework for measurement and schooling policy document, for example, which includes not a single reference to teachers, means that both the strategy and implementation of this policy has neglected its greatest resource. However, a new system that includes continuous classroom-based assessment, and builds trust and capacity within the teaching profession, holds greater promise.

7 A new national assessment system

In this report we suggest that rather than try to fix some of the evident problems that current NAPLAN has, Australia needs a new national assessment system. More detailed analysis is in Part 2 of this report. The following 5 recommendations form the basis of this new national assessment system:

1. Scrap the existing census-based testing in years 3, 5, 7 and 9 for the following reasons:
 - a) the current system is being used for the purposes it has not been designed to serve.
 - b) the initial design of the program was not oriented to teaching and learning and was based on outdated drivers for school improvement.
 - c) the Australian and international experience suggest students and the education system as a whole are best served through national assessment system that is based less on high-stakes census-based tests and more on sample-based and teacher-led assessments.

- d) the design and implementation of the program did not include teachers' perspectives sufficiently and many teachers are now disillusioned with NAPLAN.
 - e) current arrangements with a set date, one size fits all approach and annual high-stakes census-based testing do harm to some students.
 - f) the current system is not in the best interests of all students' learning and wellbeing.
2. Develop a new National Assessment System (NAS) to address the following purposes:
 - a) **Student learning purposes:** The Australian national assessment body should develop a validated, formative classroom-based Assessment Resource System (ARS), including a library of curated, quality assessment tasks and a national item bank. The ARS will be used to provide formative, teacher-led assessment for years 3 through to year 10, carefully linked to the curriculum, informed by the teaching profession and integrated with teaching and learning. These digital resources are to be used by teachers at the time of their choosing. The ARS should support teachers, complement other teacher-led assessment and enrich current student reports. All the ARS resources are mapped to and closely aligned with the Australian Curriculum and benchmarked to national standards.
 - b) **Parent reporting purposes:** that results from formative classroom-based assessments using the ARS can be included on student reports and made available to parents in a standardised, visual and plain English format that reports on students' performance in various areas of knowledge and skills over time.
 - c) **System monitoring purposes:** Annual national sample-based testing of students should be conducted for years 4, 6, 8 and 10. Curriculum areas to be included in these assessments should be diverse and alternate from year to year according to 4-year national assessment plans.
 3. That a *quadrennial program of monitoring and evaluation be developed* to ensure the national assessment system is adaptive, running optimally and not causing any unintended, negative consequences for Australian education. Monitoring will include student performance on assessments as well as questionnaires and feedback from all assessment stakeholder groups in order to develop a greater understanding of the broader aspects of schooling and factors affecting student learning.
 4. That substantial teacher input into the development of the system be undertaken, with teachers contributing to the design of the frameworks as well as assessment resources for the ARS library and item bank. A successful national assessment system that delivers improved student academic and wellbeing outcomes will only occur if there is a strong *partnership with the teaching profession*.
 5. That a new *program of professional learning and development for teachers* be developed alongside the new national assessment framework. This will have a focus on the development of professional skills with a focus on assessment literacy and assessment for learning, and efficient and productive use utilisation of the ARS.

The need for change is driven by the importance of having an assessment system that maximises the potential for student learning without the risk or harm of any kind to them. These might seem very obvious aims for any aspect of an educational system, but they have not been the focus of the current National Assessment Program, nor the many reviews that have examined it. Reorienting Australian educational assessment to students, teachers, parents and schools, as well as system monitoring requirements, provides the prospect of moving on from the unintended consequences, negativity, conflicts and tribulations now associated with NAPLAN; and ushering in a new era in Australian education.

PART 2: PUTTING NAPLAN TO THE TEST

1 NAPLAN: Contexts and background

The National Assessment Program Literacy and Numeracy (NAPLAN) plays a prominent role in the Australian education system. Over the last 12 years NAPLAN has received more public and media attention than any other education issue. This report provides an analysis of the issues surrounding NAPLAN.

We begin this analysis with the most important question of all: what is in the best interests of students? We orient all our thinking in this report to this question because it should shape the answer to every question about every element of the Australian education system. There are strong arguments about the need for NAPLAN data for system evaluation, teacher accountability and school accountability. These are all important considerations, which should contribute to the machinery of a healthy education system. However, a full evaluation of NAPLAN must include a substantive focus on how the assessment functions from the perspective of students. Students are the primary stakeholders within the assessment system and their needs should be of primary consideration. It is perhaps surprising that until now this has not been the case.

When reviewing a program like NAPLAN it is usual to assess the program against its stated aims. However, as this report will later explain, the aims for NAPLAN were not particularly clear at the outset; and have since been debated, changed and complicated through the course of its ten-year history. They have also been focused primarily upon system level considerations and issues related to students and their parents come lower on the list and teachers, for some reason, are rarely mentioned¹.

The recent review of NAPLAN commissioned by the New South Wales, Victorian, Queensland and ACT governments, claims to have identified five aims that (McGaw et al., 2020):

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

It is not possible to source these aims in a single document and a thorough review of government documentation and media releases finds a complicated picture. Initial documentation of NAPLAN's aims was not clear, and variations and additional aims added over time (e.g., NAPLAN to be used as diagnostic assessment). This lack of a clearly expressed set of aims makes any evaluation of NAPLAN against those aims very difficult. What is more, the order of the points, and language used, convey a sense that system considerations (monitoring

progress, accountability, performance and improvement) sit above those of students and parents whose needs are only covered in relation to 'achievement', 'growth', 'information' and 'performance'. The needs of students, parents and schools go far beyond these and a national assessment system should also seek to inspire 'confidence' and promote holistic conceptions of education and wellbeing. We find the above mentioned five points in the McGaw et al. (2020) review an unsatisfactory framework for evaluation.

By contrast, this report takes a different tack in assessing the relevance and effectiveness of NAPLAN and its ongoing contribution to education in Australia. Instead of evaluating NAPLAN against the originally stated, but confusing and subsequently amended aims, this report seeks to identify whether NAPLAN, in its current form, has any ongoing net benefit to students, parents, teachers and schools. An important aspect of this evaluation is considering any unanticipated detrimental outcomes. Educational assessment should first and foremost do no harm, but because NAPLAN has been subject to many aims, and claims, it has frequently been used for purposes beyond that which it was designed for, often with negative outcomes. These are not unusual dynamics in social policy systems, like education, but after 10 years, they must be assessed and attended to.

It is crucial for any education system to be prepared to evaluate its component parts and to reform or jettison features if they no longer perform a useful purpose in the contemporary context. It is not sufficient to claim that any feature of an education system should remain in place simply because it exists and has existed for some time. Therefore, in addition to prioritising consideration of students, parents and teachers, we explore and evaluate NAPLAN by considering how successfully it addresses nine key challenges facing all educational assessment systems around the world.

These challenges provide a contemporary and forward-thinking framework for evaluation of NAPLAN. The first seven of these challenges were identified in the OECD (2013) review of international practice in educational assessment and evaluation; with an additional eighth challenge added to recognise the central role of teachers in assessment.

The challenges are:

1. aligning educational standards and student assessment
2. finding a balance between summative and formative assessment
3. balancing external assessments and teacher-based assessments in the assessment of learning
4. developing fair assessments to all student groups
5. designing large-scale assessments that are instructionally useful
6. ensuring fairness in assessment and marking across schools
7. securing informative reporting of student assessment results
8. ensuring the assessment is informed, valued and of optimal utility to the teaching profession.

We should acknowledge at the outset that the eight challenges for assessment that we use to explore issues with NAPLAN, are in contrast with some of the original aims proposed for

NAPLAN in Australia and certainly those outlined by the recent review which include national monitoring of system performance and accountability (McGaw et al., 2020).

These eight challenges frame eight sections of this report. Following discussion of the issues as to how NAPLAN measures up to these challenges, this report also provides a valuable international perspective on assessment systems. Three international education experts have outlined how Singapore, Ontario, Scotland and Finland have seen their education systems evolve over the last decade particularly with respect to the use of national assessment standardised assessment tools.

NAPLAN was born within the era of the Global Education Reform Movement (GERM) (Sahlberg, 2016). Sahlberg identified key features of this movement, including “the growth of high stakes accountability and the use of corporate management practices as the key features of the new orthodoxy”. Drawing on a range of scholarly work, Acosta and colleagues (2020) define high-stakes testing as “an assessment model where students’ achievement scores on large-scale, state-wide tests serve as a criterion for making decisions about student academic growth, teacher instructional effectiveness, and school performance (Acosta et al., 2020). Clearly NAPLAN conforms to this definition.

Over the past decade there has been increasing realisation that the promises of this high-stakes accountability movement have been hollow. Concurrent to this realisation, research on educational assessment unveiled, on the one hand, the hazards and unintended consequences of high stakes assessment (von der Embse & Witmer, 2014; Farvis & Hay, 2020; Acosta et al., 2020) and on the other hand, the untapped potential of assessment to drive learning (Swaffield, 2008). The OECD criteria reflect this shift in thinking; and acknowledge the potential of assessment systems to not only operate as accountability frameworks, but also contribute to positive teaching and learning cultures and outcomes. This shift, as readers will see, has already led to substantial shifts in educational systems internationally; and has been instrumental in the assessment policies and reforms internationally including in Finland, Ontario, Singapore and Scotland summarised in this report.

Our analyses of NAPLAN and the accounts of change in other countries prepare the way for a new national assessment system that is described in the first part of this report. The analysis considers reforms that can be made to Australia’s national assessment system in order to better address the needs of students, teachers and parents, meet the challenges identified by the OECD, optimise the potential of assessment for learning as well as accountability and address the goals of Australian education. In this report we engage in an exercise to answer the question: “*what would a national assessment system in Australia look like if it were being designed today?*”

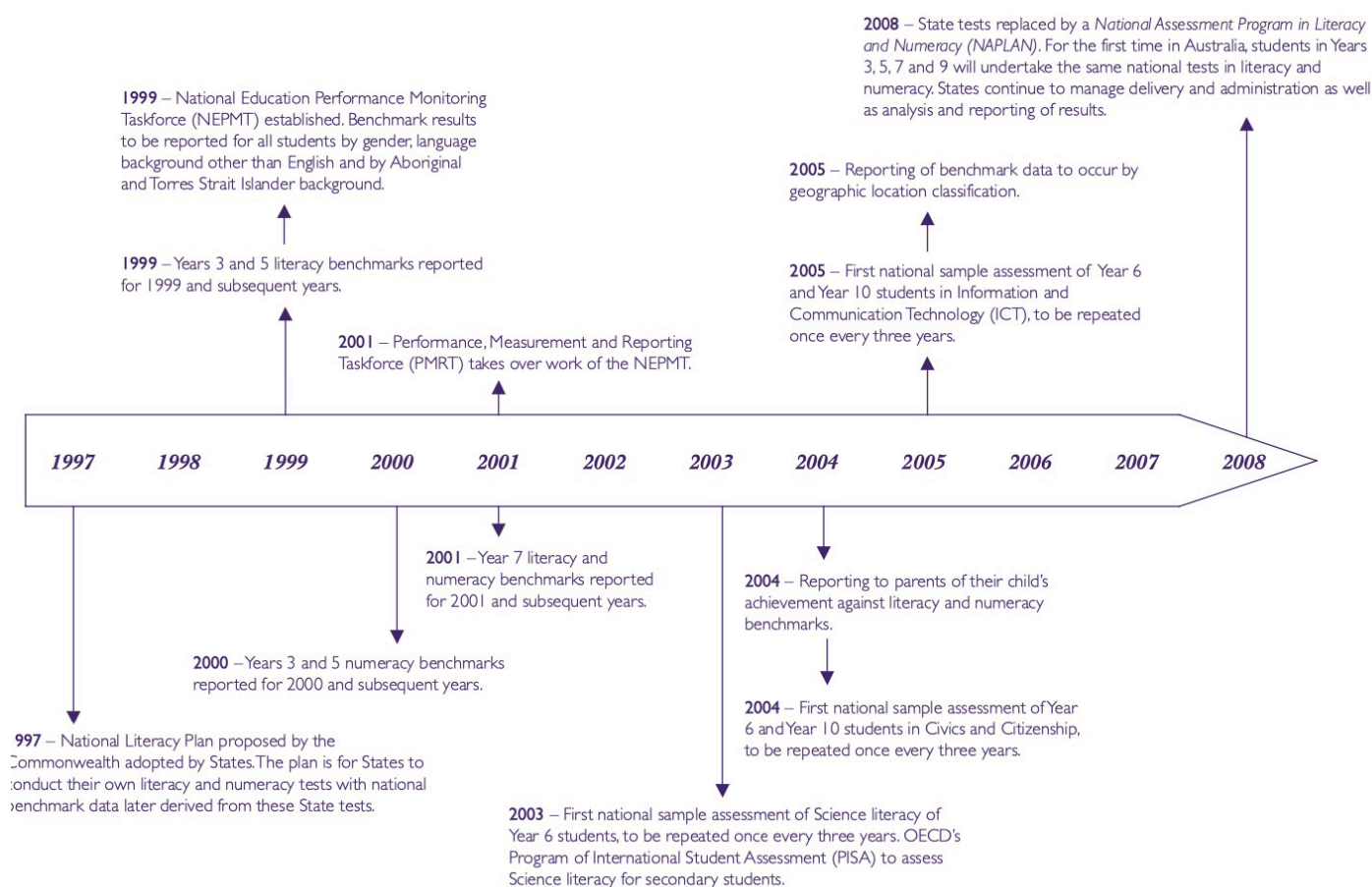
Australia’s national assessment plan for education was conceived in 1997 by the federal Howard Government, with all states and territories eventually agreeing to abandon their existing forms of testing and shift to national, large-scale standardised testing and reporting.

The National Assessment Program (NAP) includes annual, full-population (or census-based) testing of students in Years 3, 5, 7 and 9 known as the National Assessment Program Literacy and Numeracy (NAPLAN) and also staggered triennial, sample-based testing in science literacy (NAP-SL, Year 6 students), civics and citizenship and ICT (NAP-CC, NAP-ICT, Years 6 and 10).

Our report focuses specifically upon NAPLAN, which includes a series of tests in literacy (reading, spelling, grammar and punctuation since 2008, and writing since 2011) and a single numeracy test (since 2008). Since 2010 the Australian Curriculum, Reporting and Assessment Authority (ACARA) has been responsible for the NAP and also the national Australian Curriculum (AC). Testing takes place annually towards the middle of the school year, in May, with the reporting of results in August or September, depending on state or territory test administration authority (ACARA, 2020).

Prior to the NAP many states and territories had well established literacy and numeracy assessments that were well received by teachers because of the valuable data and professional support that accompanied them (e.g., NSW Basic Skills tests). Expectations for NAPLAN were set by these relatively successful tests. The NAP program evolved to avoid the complexities needed to equate state and territory assessments and assess performance against national benchmarks. Figure 4 presents a timeline of the early developments.

Figure 4. Timeline for lead up to NAPLAN implementation



(source: Dowling, 2008)

The first NAPLAN tests were developed in 2007 and administered under the federal Rudd Government in 2008. The first NAPLAN report produced by ACARA (2008) acknowledges this significant achievement:

By locating all students on a single national scale, which maps the skills and understandings assessed, each scale provides significantly more information about the literacy and numeracy achievement of students than was previously available.

NAPLAN is benchmarked against ten performance bands, which cover the spectrum of achievement from the beginning of Year 3 to the end of Year 9. Reporting on each students' performance is referenced to a range of six bands that are relevant to their school year grouping. National minimum benchmarks for each school year have been identified and the percentage of students not meeting these benchmarks is reported at school, state and national level. The logic is that clear identification of these students enables appropriate policy and practice to lift their attainment, however research suggests this is not happening. Using NAPLAN data, Adams et al. (2020) found that identification of students below the national minimum standards had no significant effects on later reading or numeracy achievement.

It is important to note that NAPLAN was introduced before the implementation of the Australian Curriculum (AC), which was put in place across all states and territories in 2015. The NAP preceded even the earliest conceptualisation of the curriculum, with the first version of the outline paper *The Shape of the Australian Curriculum* published in 2009, after testing had started. Developing the test items without the underpinning of a common curriculum caused some initial doubts about NAPLAN's validity. When states and territories adopted and adapted the AC, on differing timescales, concerns about NAPLAN's validity across jurisdictions persisted. It wasn't until 2016 that NAPLAN was directly aligned with the skills and content in the Australian Curriculum English and Mathematics:

Since 2016, NAPLAN testing has been aligned to the Australian Curriculum: English and the Australian Curriculum: Mathematics. The Australian Curriculum for English and Mathematics incorporates the essential learnings described in the Statements of learning for English and mathematics, on which NAPLAN tests were based prior to 2016. It provides the common curriculum for teaching and learning in English and mathematics for all Australian students. ([ACARA](#))

As many teachers assumed that the tests had been fully aligned to their curriculum prior to this, and with rumours regarding a shift in content, ACARA clarified the situation in a [letter to all Australian teachers](#).

Three further events influenced the evolution of NAPLAN. First, the introduction of the [My School website](#) by the Gillard Government in 2010. Second, the phased introduction of [NAPLAN online](#). Third, the restriction of access to NAPLAN test past-papers in 2016. These developments had significant impact on the nature and purpose of NAPLAN.

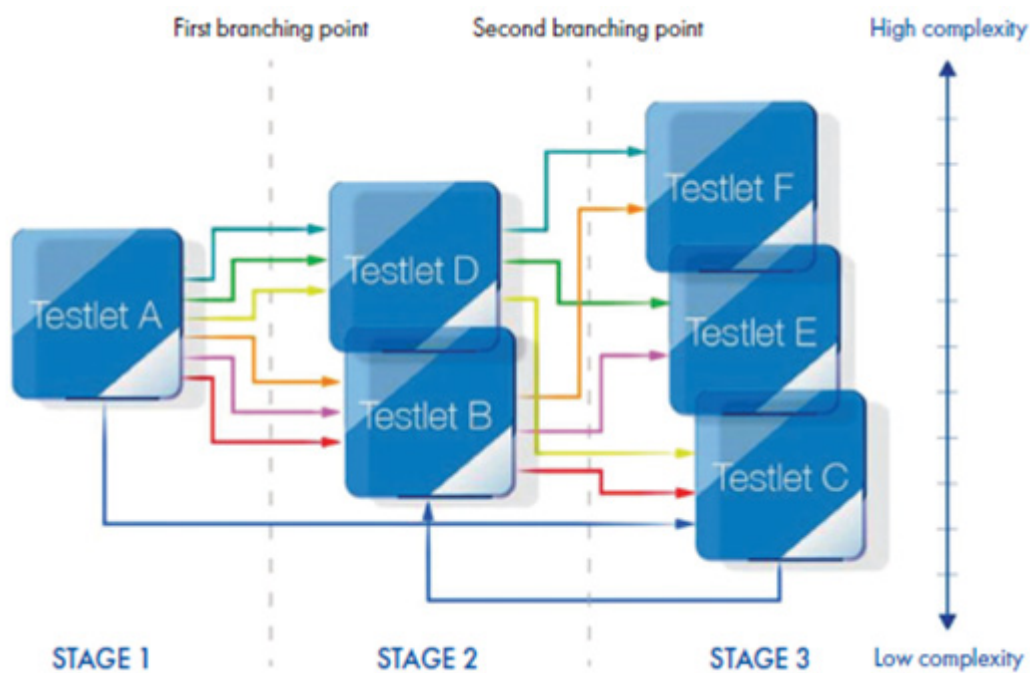
[MySchool](#) is a publicly accessible website and database "resource for parents, educators and the community to find important information about each of Australia's schools". The stated purpose of *MySchool* (2019) website is to provide:

...information that allows everyone to learn more about Australian schools, and for schools to learn more from each other. For parents, My School provides information to help make informed decisions about their child's education.

The linking of NAPLAN and *MySchool* website, a resource at the centre of school choice policies, pivoted the purpose of the tests away from a system level monitoring approach and towards a school performance league table. This development attracted a great deal of media and public attention (Mockler, 2013).

Similarly, current plans and the initial transition to NAPLAN online testing, attracted extensive media attention. Online assessment enables adaptive testing, designed to produce a test that is more responsive to the needs and abilities of students. NAPLAN adaptive testing has been designed using testlets (Figure 5). Student performance on testlet A determines which testlet they are directed to next, and so forth. This ensures that in Stage 2 and 3 students are presented with testlet items that are most suited to their attainments in the preceding stages.

Figure 5. Online NAPLAN design using testlets (source: Adams, 2020)



NAPLAN online testing, incorporating these computer adaptive testing methodologies, was trialled in 2017 and a phased transition to online began in 2018. ACARA aims for all tests to be fully online by 2022. The early transition to online testing was criticised in terms of preparedness and implementation, as many schools experienced technical difficulties. In 2019 many online tests malfunctioned, and students had to sit them again (Baker and Cook, 2019). Many also questioned whether it was necessary, or appropriate, to shift away from handwriting to computer word processing. Here again, alignment with the enacted curriculum in schools was questioned, particularly in relation to young children in Year 3, who are novices in computer navigation and word processing, and most classroom work is done using pen and paper. A timeline of the development and planning for online NAP is available in Annex 2.

From the beginning in 2008 NAPLAN past test papers and answers were made available on the [ACARA website](#). This was in line with the often-flaunted idea that NAPLAN, as an open and transparent testing system, could be used to promote teaching and learning. However this practice ceased in 2016 and a [subsequent “freedom of information’ request](#) was rejected by the National Information Commissioner (NIC). ACARA contended that the release of tests jeopardised development of the online tests, which required more items from past papers to equate the annual tests within the adaptive testlet system. The NIC decision notes that protocols outline:

Principals (and their teaching staff) are not permitted to provide NAPLAN tests to parents/carers or members of the wider community, including the media, even after the test security period has ended.

Regardless of the strength of the rationale for limiting access to test papers, the development has again shifted the orientation of NAPLAN and reduced transparency. The decision has stopped public and media scrutiny of test content; and contributes to the perception that these are high-stakes tests. Although this has limited access by students and families, it does not appear to have dented the proliferation of commercial practice tests and NAPLAN preparation products. Unlike other issues with NAPLAN, this one has received very little media coverage; but it supports the view held by many teachers, that NAPLAN tests are high-stakes and focused on accountability and compliance, rather than teaching and learning (McGrath-Champ et al., 2018).

2 How does NAPLAN meet the assessment challenges?

Nothing is perfect in the world of education as countries continually strive for lofty, and sometimes contested and shifting, educational ideals in a rapidly changing world. National assessment systems are not perfect and are often the focus of particularly vociferous critique. All national assessment systems face common challenges. The 2013 OECD [review of international practice in educational assessment and evaluation](#), identified seven key challenges for student assessment systems (OECD, 2013):

1. Aligning educational standards and student assessment
2. Finding a balance between summative and formative assessment
3. Balancing external assessments and teacher-based assessments in the assessment of learning
4. Developing fair assessments to all student groups
5. Designing large-scale assessments that are instructionally useful
6. Ensuring fairness in assessment and marking across schools
7. Securing informative reporting of student assessment results

For the purposes of this report, we have added a further challenge for Australia’s assessment system:

8. Ensuring the assessment is informed, valued and of optimal utility to the teaching profession

In the context of school education, effective assessment requires both design for Assessment for Learning (AfL) and professional implementation of it within classrooms. Historically national assessments and teacher classroom assessments have been dichotomised, with the view that the former is for determining system learning outcomes and enabling analysis of related factors and responsive policy while the latter is for improving teaching and learning in classrooms and schools. More recently however, even some of the largest scale national assessment systems have aimed, and been designed specifically, to meet both accountability and pedagogical goals. To complement other reviews of NAPLAN, which have focused primarily on how it is fit for system reporting and accountability purposes, our assessment of NAPLAN considers pedagogical goals and how NAPLAN is aligned with, or otherwise effects, learning. Recent research has highlighted both the difficulties and promise of bridging the gap between assessment and learning. Baird and colleagues (2017) argue that:

If assessment is to serve the learning goals of education, then ... discussion on the relationship between assessment and learning should be developed further and be at the forefront of high-stakes, large-scale educational assessments.

Points 2, 3, 5 and 8 all outline challenges for assessments, like NAPLAN, to make a positive contribution to teaching and learning. While NAPLAN aims inferred the system would drive school improvement, neglect of its relationship with teaching and learning, and inadequate attention to the views and experiences of teachers, has meant that potential has not been fully realised.

Recent research has highlighted the pivotal importance of teacher professional engagement with national assessment programs (Heitink et al., 2016). Teacher engagement is necessary in order to meet many of the assessment challenges listed, as teachers play a key role in the implementation assessment process. In the case of NAPLAN partnering with teachers is an essential element of 'primary stakeholder engagement' necessary for high-fidelity implementation of any large reform (Salloum et al., 2017); but an aspect not widely considered during the introduction of NAPLAN. If large-scale national assessment is to be more integrated with teaching and learning, then engagement of teachers is especially critical. Thus, efforts to address challenge 8 are likely to strengthen positive outcomes in a reformed national assessment system.

Summative and formative assessments

Several of the criteria refer to summative and/or formative assessment. These are important concepts in education and reflect the different ways assessments can be used. This dichotomy reflects the polar ends of a spectrum of uses for assessment, at one end assessment are summative and used for “summing up” what has been learnt; at the other end the assessments are formative and used to “form” learning. In other words (Dixon & Worrell, 2016):

Formative assessment involves gathering data for improving student learning, whereas summative assessment uses data to assess about how much a student knows or has retained at the completion of a learning sequence.

The dominance of accountability purposes in the NAPLAN means that the tests are oriented to summative assessment, but as we will show later, there have also been claims that they can be used formatively.

Sample- or census-based testing?

One of the key considerations in designing national assessment systems is whether the assessments and tests should be completed by all students, referred to as census-based testing, or by a small representative proportion of students, referred to as sample-based testing.

Although ‘sample-based vs. census-based’ student assessment design has been presented as a dichotomy in the past, with new technology those divisions may fade and it is possible that hybrid, flexible systems can be designed to provide benefits of learning for all – and sampling for system analysis can be undertaken as needed. Such systems might be able to draw on national assessment resources (assessment libraries, item banks and online testing systems) in a range of ways, enabling a suite of approaches for a national assessment system. This would enable sample-based assessments to be done under controlled test conditions for accountability purposes but also enable more routine and flexible use of resources for formative census-based assessment and summative census-based testing when it is appropriate.

There are many issues here to consider. *Assessment for Learning* approaches, for example, which seek to strengthen the relationship between students’ assessments and their learning, should be provided to the whole population (census) as it would be unethical to reserve the learning benefits for the few within samples. If all, or part, of a national assessment is designed to strengthen learning, all students should be able to benefit from that aspect of assessment design.

The current NAPLAN system uses annual census-based testing, although participation rates have been declining and many students are excluded from the tests. Internationally there has been a recent shift from such approaches. This is explored later in the international case studies. The possibility of developing an assessment approach that, like Scotland, utilises a hybrid approach is a central consideration in the following ‘concept paper’ report, which focuses on redesign of Australia’s national assessment system. To inform a future focussed design we first evaluate the NAPLAN system. The nine challenges provide a framework for analysis and in this report, we ask: “*How does NAPLAN measure up against these nine challenges?*”

In doing so the issues, strengths and weaknesses of NAPLAN are systematically explored and clarified.

Challenge 1: Aligning educational standards and student assessment

For many countries this challenge relates to the linking both curriculum and assessment to standards. The fundamental assumption that school assessment is aligned to school curriculum has been established for many years (Tyler, 1949), however in recent decades this has been enriched by the understanding that alignment of assessment to *written standards* can be used to further drive teaching and learning (Drake, 2007). However, in the case of NAPLAN the fundamental assumption that the tests are aligned to curriculum is a point of weakness.

While NAPLAN was designed with alignment to standards written in the 10 Bands, much of NAPLAN critique focuses on NAPLAN's weak alignment with curriculum. There are two main issues here:

1. Poor early alignment between NAPLAN and the taught curriculum, and ongoing confusion, particularly among teachers.
2. Narrowing of the taught curriculum, due to the privileging of NAPLAN over other NAP and non-assessed curriculum areas.

As noted in the brief review of the NAP history, NAPLAN was not originally aligned to the Australian Curriculum, where standards in curriculum content and skills are implicit. Some experts assert that NAPLAN was never designed to align with curriculum per se, as NAPLAN assesses the broader constructs of literacy and numeracy (Tognolini, personal communication, 2020). However, early statements from ACARA claim that prior to 2016 the tests were aligned to the 2005 [Statements of learning for English](#) and the 2006 [Statements of learning for mathematics](#). From 2016 onwards the tests were aligned to [the Australian Curriculum](#). Currently the Australian Curriculum provides "learning continua" for literacy and numeracy and alignment with these is assumed but has not been researched or evaluated. Certainly, unless told otherwise, students, teachers and parents would assume that a national assessment is strongly aligned to school curriculum.

From the outset though, many teachers were confused about the alignment with curriculum and standards; particularly as many were teaching solely from state curriculum for up to the first four years. Since the national transition to the Australian Curriculum, in 2015, teachers continue to deliver the AC via customised state syllabi – some of which have substantial curriculum beyond the AC. Teachers' confusion is evident in a wealth of correspondence and media reports, and was responded to in a [letter from the CEO of ACARA to all teachers in 2016](#). The letter explained the previous alignment and the transition to alignment with the Australian Curriculum, but it also introduced additional complexities. These are again, compounded by the delivery of the Australian Curriculum via state and territory frameworks, and overall, the situation is complex and difficult for teachers (and also students) to understand. The alignment is also difficult for parents to understand. A study (Colmar-Brunton, 2016) commissioned by ACARA found:

Parents' knowledge of the purpose and logistics of NAPLAN is often limited. There is a lack of clarity around what is actually tested and a range of views around how results are used.

Over the years both parents and teachers have been further perplexed by the intents and purposes of NAPLAN, including contradictory claims that NAPLAN is aligned to the Australian

Curriculum but that it cannot be taught to, as former ACARA chairman Barry McGaw contended in 2011.

How NAPLAN assesses against the AC, which has been acknowledged as large and crowded (Donnelly & Wiltshire, 2014), while focusing on four points (Years 3, 5, 7 and 9) is not easily understood. ACARA does state that much of the tests are based on the previous school year; but in total there is substantial complexity in the alignment between tests, the test standards and the curriculum.

There is no doubt that stronger and more temporal alignment between the tests and curriculum is possible. Such alignment provides the theoretical foundation for hope that the tests result in productive teaching and learning of the prescribed curriculum in classrooms. Logically, poor alignment, particularly when in tandem with high stakes and performance pressures, would lead to *teaching to the test* that extends beyond the normal school curriculum; and introduces new time and performance pressures upon teachers and students. Numerous studies suggest that this is precisely what has happened with NAPLAN (Lingard & Sellar, 2013; Wyn, Turnbull & Grimshaw, 2014; Howell, 2015; Cumming et al., 2018; Lingard et al., 2016).

Additional difficulties, and confusion, are precipitated by the fact that since 2016 the tests themselves are no longer available to the public, nor indeed to professional educationalists or journalists (Australian Information Commissioner, 2015). Whilst a large market in commercial NAPLAN products and services has developed (Ragusa & Bousfield, 2017), with practice tests, we are now in the situation where students and parents are [not able to access the recent test papers](#) and in schools tests are managed under [strict security protocols](#). This may be a symptom of the fact that the tests are now *high-stakes*². In addition, it hinders transparency and understanding of how tests are aligned to standards and curriculum.

While alignment and transparency are substantial issues, the issue that has attracted the most attention is the privileging of NAPLAN tests within the schooling system and the narrowing of curriculum focus that it has caused. The commentary around NAPLAN, in media and research, asserts that this privileging is linked to the inclusion of NAPLAN data on the *MySchool* website and the associated performance pressures this imposed upon schools, teachers and students (Comber, 2012). The publication of results shifted the focus of NAPLAN from monitoring and enhancement of student learning to competition among teachers and school leaders (Thompson, 2015).

Although NAPLAN was not originally designed as a high-stakes assessment, substantial evidence suggests that linking it with *MySchool* website has made it so (Lobascher, 2011; Dulfer et al., 2012). Research by Bob Lingard and Sam Sellar (2013) explored how in the early years of NAPLAN, data was used to set performance targets in literacy and numeracy for states in return for monetary rewards thus attaching 'high stakes' pressure to the performance of states, which cascaded down to regions, schools and classrooms.

The high stakes nature of NAPLAN has also been acknowledged in recent reviews of NAPLAN (Cummings et al, 2018; Loudon, 2019), although in some cases this is notably demoted as merely "stakeholder concerns" (Masters et al., 2020). A large body of research documents the negative effects of High Stakes Testing (Harlen et al., 2002; Acosta et al., 2020) and recent research explores how it is considered to lead to distortion and narrowing of curriculum in schools (Farvis & Hay, 2020).

Research and commentary on NAPLAN consistently assert that teachers and parents believe it has distorted the curriculum, teaching and learning (Dulfer et al., 2012; Carter et al., 2018; Matters, 2018; Cummings et al., 2018). This has been countered by recent claims that over 80% of school principals do not agree that the curriculum has been narrowed, or that too much time is spent preparing for the tests (Louden, 2019; Adams, 2020). This contradiction is perplexing, and on balance it seems that the teachers' report is likely to be more closely supported by direct personal experience, hours in classrooms and direct observation. Teachers' views are also supported by logic, and international research, which makes it clear that when assessment becomes high stakes more teaching time is devoted to teaching to the test and that may detract from time spent on other teaching (Wyn et al., 2014). In a recent study of 109 education consultants in New York, Farvis and Hay (2020) received consistent reports that high stakes testing had led to:

reduced teacher control in instructional planning, curriculum narrowing and increased test preparation. These practices were associated with decreased collaboration and increased teacher and administrator stress.

Many argue the heavy focus on NAPLAN is unhealthy. More than 60% of parents believe that too much time is spent preparing students for the test (Matters, 2018). Milligan reminds us that while literacy and numeracy are important "these outcomes and indeed outcomes of any kind - are a very long way from being the only fundamental things in schooling" (2019). Early research surveyed over 8,300 teachers from every state and territory, shortly after the introduction of *MySchool* website, and showed:

- narrowing of teaching strategies and of the curriculum
- negative impacts on staff morale, and
- negative impacts on some schools' reputation and capacity to attract and retain students and staff (Dulfer et al., 2012).

More recent studies have made similar findings. A survey of English teachers' views on NAPLAN Carter, Manual and Dutton (2018) found that, even when focusing on English as a Key Learning Area (KLA):

- The tests added little to teachers' understanding of literacy levels
- The assessment was a poor and narrow measure of student achievement
- Teachers felt pressured to prepare students for the tests and this detracted from other learning opportunities
- The pressure to "teach to the test" frustrated many teachers and reduced their sense of professional autonomy

Overall, the full national assessment programme has a less than comprehensive coverage of KLAs as testing covers only literacy, numeracy, and sample-based tests in science, ICT and Civics and Citizenship. Students, parents and teachers are highly aware that whatever is taught in the valuable, and often treasured, fields of history, geography, physical education, languages, music and art is not assessed. The census-based testing and high stakes attached to NAPLAN favour and privilege literacy and numeracy over other subjects many of which are critically important to 21st century life and work, such as technology and science which are not assessed

across the full population and problem solving, and critical thinking which are not the focus of any of these assessments.

Evidence that NAPLAN has narrowed curriculum focus is now so substantial that even state and territory Ministers for Education have publicly voiced their frustrations and suggested that major reform is needed. NSW Minister Rob Stokes noted in 2018:

You now have an industry that's grown up alongside it, where teachers are being encouraged to teach to the test rather than the curriculum...it's time we had discussions about replacing it" (Baker, 2018).

In conclusion it is clear that, although there have been efforts to align NAPLAN to curriculum standards across all states and territories, the picture is complex and alignment to curriculum, particularly in relation to timing, could be tighter. Further, the lack of alignment, poor communication and transparency, and the high stakes attached to NAPLAN, have all frustrated teachers and resulted in a narrowing of curriculum focus on classrooms.

Challenge 2: Balance between summative and formative assessments

Three key points in relation to the summative and formative³ balance in NAPLAN are:

1. The original design of NAPLAN was as a *summative assessment*. There is no evidence that NAPLAN has produced positive outcomes as a formative assessment. Data suggests the contrary, but there is potential for the summative data to be used for more formative purposes, particularly via online adaptive testing.
2. There have been *mixed messages from political and educational authorities on the purposes of NAPLAN*, including its potential for summative and formative assessment. These have led to confusion and frustration.
3. There is evidence that NAPLAN has had negative impact on teaching and learning. Overall, *NAPLAN could have stronger positive alignment with teaching and learning*, and more formative orientation to its assessments.

NAPLAN was primarily designed as a system level summative performance measure, but from early on there was also mention of "the diagnostic value" of its assessments, which might contribute to formative assessment.

Initially NAPLAN was linked to performance and accountability frameworks for schools. After the launch of the first tests in May 2008, Education Minister Julia Gillard made her views on the purpose of NAPLAN and the *MySchool* website clear in a media statement in August 2008.

We can learn from [Joel] Klein's methodology of comparing like-schools with like-schools and then measuring the differences in school results in order to spread best practice," she said...The answer is not league tables and it's not A to F reporting, but it's making sure we have this rich performance information available, school by school. (Ferrari and Bitá, *The Australian*, August 12, 2008)

In 2010, then Education Minister, Julia Gillard (2010) wrote a [letter to the President of the Australian Education Union](#) making it clear that the Government was committed to:

1. the *transformative power* of high-quality education
2. the value of *transparency and accountability for ongoing school improvement*
3. effective assessment and reporting of student achievement including the *diagnostic value of the NAPLAN*
4. the power of accurate data to provide *comparable national reporting of school effectiveness*
5. *ensuring that educational disadvantage is addressed*, and an opposition to the misuse of student performance data including simplistic league tables.

An assumption sitting behind these points and many of the subsequently stated purposes, was the belief that NAPLAN would lift national educational attainment. Despite widespread reports that teachers are *teaching to the test* and NAPLAN is the focus of a great deal of classroom work, the expected improvement has not materialised. Some gains have been noted over the decade, particularly in some states and student groupings, however nationally the magnitude of improvements is small and many of the tests show no gains.

Since 2008 statistically significant but small gains are evident only in Years 5 and 9 numeracy, Years 3 and 5 reading, Years 3 and 5 spelling, and Years 3 and 7 grammar. Thus, many year groups and test domains show no significant progress. Significant and consistent declines are evident in writing test results across Years 5, 7 and 9 since the first year writing was assessed in 2011. NAPLAN serves its summative purpose by identifying and monitoring these results but a more formative assessment, specifically designed to improve teaching and learning, may provide better outcomes.

Over the decade, there have been many claims that NAPLAN is 'diagnostic' and can be used in a formative way. As the tests assess students against the same standards over the course of their schooling, there is potential for the tests to be formative and many initiatives have tried to maximise the formative aspect by focusing on how teachers' can use NAPLAN data ([Victorian Department of Education, 2013](#)). However, many of these efforts have frustrated teachers because the design of the tests and timing (testing in May reporting in August/September) means that using the testing process and NAPLAN data to improve teaching and learning is just plain difficult (Renshaw et al., 2013). Despite this obvious weakness, the diagnostic and formative potential of the test has continued to be spruiked; and teachers, parents and students have been bombarded and confused by contradictory messages:

From the ACARA 2008 National Report on schooling (ACARA, 2009):

Importantly, NAPLAN can be used by teachers for diagnostic purposes.

ACARA Chief Executive Peter Hill (2010):

"The purpose of national testing has been to get a snapshot of student performance for reporting back at different levels. That was the purpose from the beginning, and the purpose has never been diagnostic assessment."

ACARA Chair Barry McGaw (2011) said:

“NAPLAN is not a test students can prepare for because it is not a test of content. The federal government’s intention in introducing and reporting NAPLAN results was to provide a diagnostic tool for teachers and parents, identifying gaps in students’ skills.”

The 2013 Senate Enquiry into NAPLAN noted:

“It is apparent in the submissions to this inquiry that there is confusion and inconsistent statements about the purpose of NAPLAN.”

Refuting the diagnostic and formative potential, ACARA CEO Rob Randall (2014) said:

“NAPLAN is not in itself a means of improving the quality of education, but it gives us data to have objective discussion about what is working and what is not.”

Sarah Mitchell, NSW Minister for education (Baker & Cook, 2019):

“[We’re] supportive of standardised testing, but also looking at formative testing” and “In 2019, it is clear that a diagnostic test must be on demand, it must be linked to the curriculum, it must focus on student growth and it must test informative writing. NAPLAN in its current form does not meet this criteria.”

With such contradictory statements coming from a wide range of political and educational authorities, teachers have been left confused. Teachers have been subjected to rhetoric on use of NAPLAN as formative, whilst facing practical barriers to realising formative potential. Many have attempted to maximise NAPLAN’s utility as a formative assessment via data-based AfL and many school data analysis projects (e.g. ‘data walls’), only to be faced with mounting frustration. While there is potential for all assessment to be used to improve teaching and learning, the design of NAPLAN, and in particular its timing, makes this difficult and teachers perceive the tests as cumbersome, and designed for accountability rather than for teaching and learning (McGrath-Champ et al., 2018).

Acknowledging this weakness, the former General Manager of Assessment and Reporting (Adams, 2020) at ACARA recently recognised the need to:

facilitate the introduction of formative assessment into NAPLAN, so that it might play a much greater and a radically new role in the quality of teaching and learning in our schools.

Adams (2020) asserts that the established validity of the test, vast item bank developed through NAPLAN, and the recent shift to online, adaptive testing, will contribute to tailored testing that is formative. However, retro-fitting NAPLAN as formative assessment presents risks. As highly esteemed international educational assessment and testing researchers James Popham and David Berliner (2014) note:

In educational practice and policy contexts, the close inter-dependencies between test design and a test’s purposes are not always clear to all concerned players... In such high-stakes policy contexts, test misuse can

result in serious and sometimes negative consequences for stakeholders at schools.

Another difficulty with Adams' proposition is that the mixed messages of successive governments and ACARA have left teachers with little faith in the tests and seriously disillusioned. It could be argued that for many the NAPLAN brand has been stained, and efforts to revise the test to be more formative may be stymied by the brand distrust that now exists.

Challenge 3: Balance between external testing and teacher-led assessments

Although balance with teacher-based assessment is an important consideration for national assessment systems, it has received little attention in Australia. Key points to consider here include:

1. NAPLAN has not been presented as an assessment to be balanced, or countered, by strong teacher-led assessment.
2. Australia's NAP strategy documents make scarce mention of teacher led assessment.
3. It could be argued that the primacy of NAPLAN within Australian schools, has threatened teacher-based assessment.

As a high-stakes, annual test, conducted at a single point of time within the school year, NAPLAN imposes an impact that reverberates throughout the school year. It has been widely argued that test-based accountability, like NAPLAN, does not improve the quality of education (O'Neill, 2013; Sahlberg, 2010) but rather it harms the quality, equity and viability of education (Conway & Murphy, 2013; Hargreaves, 2008). To ensure that the accountability benefits of systems, like NAPLAN, outweigh these negative elements, large-scale assessment systems should support, rather than compete with, teacher-based assessments, which have well-known positive effects on learning.

Unfortunately, in the development of NAPLAN, teacher-based assessments were not part of the conversation. None of the espoused purposes of NAPLAN highlight, or even acknowledge, the importance of teacher-based assessment. A search of the ACARA website shows no mention of "teacher-based" "classroom-based" or "teacher-led" assessment (ACARA, 2020). ACARA's recent [Measurement Framework for Schooling in Australia 2019](#) makes no mention of teacher-based assessment (in fact the term "teacher" does not appear in that document!). Although, it is more encouraging to note that National School Reform Agreement (COAG, 2019) does outline goals to "support teacher assessment of student attainment and growth against clear descriptors" and "Assisting teachers monitor individual student progress and identify student learning needs through opt-in online and on demand student learning assessment tools with links to student learning resources". Rather than supporting authentic, autonomous teacher-based assessment, these developments appear to move teacher-based assessment toward accountability frameworks.

The imbalance between external NAPLAN and teacher-based assessment was identified early (Klenowski, 2015) but has not been addressed. In an OECD review of Australia's educational assessment, key areas for improvement included (Santiago et al., 2011):

- 1.1.1.1. Establishing national strategies for strengthening the linkages [of assessment] to classroom practice
- 1.1.1.2. Maintaining the centrality of teacher-based student assessment while ensuring the diversity of assessment formats

Teacher-based assessment does occur frequently within Australian classrooms. A large survey of more than 18,000 NSW public school teachers found that more than 80 percent of public school professional staff (teachers and school leaders) doing marking at least every week ; and more than 40 percent report writing and developing their own assessment tasks every week (McGrath-Champ et al., 2018) The survey also made it clear that teachers recognise that NAPLAN is focused primarily on accountability and compliance, and that it is this agenda that drives their assessment practice (Wilson and Alonzo, forthcoming).

Rather than support teacher-based assessment, there is growing evidence that external high-stake tests, like NAPLAN, put pressure on teachers and schools to comply with an examination-driven culture to meet the accountability agenda (Conway & Murphy, 2013; Datnow et al., 2019; Sahlberg, 2010). Internationally, accountability focused assessment systems are prominent and frequently occur alongside teaching quality frameworks that use a single measurement for data, and this is perceived by teachers and principals to devalue their work and threaten their integrity (McNeil et al., 2008). There are also arguments that MySchool website has changed the way that teachers and schools are viewed, with some parents placing too much emphasis on the tests, and not enough on the other learning activities and assessments undertaken during the year. Thus NAPLAN and *MySchool* website play a central role in developing a culture of audit; feeding media discourses and quantifying what it is that education is, and should be, doing (Gannon, 2012; Mockler, 2013).

There is limited evidence, other than the OECD review, with which to address the question of 'balance' between NAPLAN and teacher-based assessment. However, in the NSW survey, teachers reported that their assessment activities were becoming mechanistic, and limited in how they could be used to improve teaching and learning. Teachers' comments included: "Even with all the results of assessments, teachers do not have time to reflect on the assessment findings anyways" and "I feel there is a lot more pressure to complete data and assessment and testing and at times teaching and coaching to the test in NAPLAN" (Wilson & Alonzo, forthcoming).

Challenge 4: Fair assessment of all students

There are many issues here:

1. Research suggests NAPLAN tests are not suited, nor fair, to substantial proportions of Australian students
2. The appropriateness of high-stakes NAPLAN testing may be different for different age groups

3. Participation rates for NAPLAN tests have been declining, suggesting a lack of suitability and engagement that may also reflect unfairness.
4. Commercialisation of NAPLAN preparation also presents a threat to equity; with more affluent families able to 'buy' support and better results.

Issues of fairness, and more specifically poor validity and reliability, have been identified in relation to:

- **Indigenous students.** examining the impact of NAPLAN on children and teachers in remote Australian Indigenous communities, researchers concluded these students are not treated equitably by the test instrument or reporting for accountability purposes (Jorgensen, 2010; Macqueen et al., 2018). The researchers suggest the tests are culturally inappropriate (Jorgensen, 2010). In particular the cultural nature of standardised testing, as an individual, timed activity that considers co-operation to be cheating and encourages competitiveness, may prevent Indigenous, and other minority, students from fully engaging (Macqueen et al., 2018).
- **Disadvantaged students.** The single sitting, high stakes nature of NAPLAN suggests that reliability and validity will be lower for lower attaining students who are likely to be from disadvantaged backgrounds (Harlen et al., 2002). The choice and competition effect of *MySchool* website is seen also to widen the achievement gap of disadvantaged children (Diamond, 2012; Wössmann et al., 2007) and the deleterious impact of national assessment upon school curriculum are felt most keenly among disadvantaged schools (Au 2008; Thompson & Harbaugh 2013). Although NAPLAN is designed to identify students not meeting minimum national benchmarks, trends show that it is not driving improvement for those disadvantaged students (Adams et al., 2020).
- **Students with disability.** Research suggests NAPLAN can be a negative experience for students with special needs because of the inadequacy of special provisions to allow these students to access and respond to the test questions in a fair way (Davies, 2012; Macqueen, et al., 2018).

Good coverage of the issues relating to fairness can be found in the 2018 Queensland review of NAPLAN (Cumming et al., 2018). Angelo (2013), eloquently explains the dynamic that occurs in some classrooms and disadvantages lower attaining students:

Practising [tests] is not teaching. Those who can do the test, who benefit by being taught the layout or format of the test and how to answer questions, can be divided from those who cannot do the test and need "teaching" to assist them to "learn" the substance. Through practice, the latter learn that they cannot successfully do the test, while those who can do the test become "test wise". Those who cannot are deprived by their class learning time being taken up with correcting the tests which they could not do. When these students participate in correcting the test, they are shown what they could not do, but not taught how to do it.

ACARA's own research suggests parents are concerned about several issues related to fairness including disadvantage for less resourced schools, stress on students, the exclusion of some students from the test, and concerns as to whether NAPLAN accurately reflects all students' performance (Colmar-Brunton, 2018). Parents also reported concerns that NAPLAN "diverts

teaching time from the curriculum”, an issue that would disproportionately disadvantage struggling students. In particular, specific critique has identified numerous weaknesses in the design of tests and marking frames which create unfairness in the Writing assessment (Perleman, 2018).

There are also issues regarding the age appropriateness of NAPLAN testing. Research analysis of the submissions of the 2010 NAPLAN Senate Inquiry suggested “*adultification* in Australian schools, with children subjected to developmentally inappropriate expectations, pressure, stress and precocious knowledge in response to NAPLAN testing and reporting” (Bousfield & Ragusa, 2013). This is a critical issue in national assessment design. The difficulties lie in the high-stakes nature of NAPLAN, which has been precipitated by publication of data on the *MySchool* website. Many countries (e.g., Finland) have opted for a progressive system of assessment which starts with classroom-based assessment, initiated at various year-levels of primary education and culminates in external, summative, high stakes assessment at end of secondary school.

NAPLAN however has been designed as a one-size-fits-all program, with similar testing arrangements for years 3 to 9, and there is no national program of assessment for senior secondary school, post 16. ACARA’s survey of parental attitudes noted that: “Acceptance develops over time, with the main concerns (if any) emerging in year three when the child first takes the test.” (Colmar-Brunton, 2016). This may be a symptom of the inappropriateness of testing at Year 3. NAPLAN also shows large Relative Age Effects; and these are largest in Year 3 (Whitely et al., 2020).

Typically, high levels of engagement, and perhaps also anxiety (as reported by parents), are seen among younger students in years 3, with peak levels of high-stakes effects at Year 5, as NAPLAN is used for competitive entry to some streams and secondary schools at this point. In NSW, for example, parents entering their child for selective school exams, are asked to provide permission to access their child’s NAPLAN results.

The tests also have declining levels of student engagement in years 7 and 9. The recent review by McGaw and colleagues (2020) found “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.” This pattern of student engagement is the reverse of what might be considered developmentally appropriate. An appropriate developmental system would expect light, pressure free, levels of engagement in the younger primary years which escalate to higher levels of engagement in senior secondary school – at an age when the majority of students have developed skills and dispositions that are prepared for higher stakes assessment and able to optimise learning for high-stakes assessment. Howell’s study (2015) of 105 primary school students’ experience of NAPLAN found that:

children experienced the tests within an emotionally charged and confusing context of contradictions and dissonances emanating from multiple sources; receiving little, if any, clear and consistent information regarding the purpose of the tests. Further, this confusing context was compounded by the need for test preparation as a result of NAPLAN’s idiosyncratic format and test protocols.

Howell's study documented numerous negative experiences among children, and noted that these were magnified among the younger, year 3, students.

Participation rates for the NAPLAN tests have declined over the decade (Anderson & Boyle, 2015; Thompson et al., 2018); and can be seen to vary according to state and territory jurisdiction and school levels of socio-educational disadvantage. The lack of participation by some student groups highlights issues of fairness in how the test is presented and implemented. Colbert, Wyatt-Smith and Cumming, (2015) outline issues relating to this and conclude that labelling and participation issues combine with negative effect:

Students who are at risk in their literacy and numeracy learning have become the lost cause in policies that emerged two decades ago to assist them. If positive engagement with schooling and long-term retention are national goals, NAPLAN may well be working against these by creating broader characterisation of risk.

A further issue is that the commercial industry surrounding NAPLAN disadvantages families with limited financial resources. Commercial products related to NAPLAN have, some would argue, developed in response to NAPLAN's transition to high-stakes assessment. It is the weight that NAPLAN carries in relation to educational concerns of parents that has developed the market for these products. There are a large variety of products available for parents to use to supplement their child's preparation for NAPLAN tests, however, they all have one thing in common, they cost money. This adds an additional level of inequity to the NAPLAN phenomena.

Despite the evident proliferation of commercial products, ACARA's CEO in 2014 corrected common perceptions that the tests were high-stakes, saying:

Also.... "Despite what some may think, it is a "low stakes" test for children, in the sense there are no rewards or punishments for the result a child achieves"

This sort of denial contributes to the mixed messages that schools, teachers, parents and students have received on NAPLAN. It refuses to acknowledge the complex ways in which the linking of NAPLAN to *MySchool* distorted the nature of the tests (Lingard & Sellar, 2013). It also refuses to recognise the now heavy commercial interests that are part of the NAPLAN phenomena and the ways in which these create relative disadvantage for many students.

Challenge 5: Large-scale assessments that are instructionally useful

The issues identified in relation to this challenge stem from some identified earlier in relation to alignment with educational standards and curriculum, and NAPLAN's weaknesses as a formative assessment. Assessments are key drivers in learning processes and there has been system adaption and accommodation to NAPLAN. However there have also been many unintentional outcomes, with issues like:

1. Teachers feel pressured to *teach to the test* due to NAPLAN's role in accountability, however, many find student preparation difficult, or challenging to their professional integrity, because:

- a. NAPLAN timing makes it less instructionally useful
 - b. Weaknesses with curriculum alignment hamper instructional practice.
 - c. Time spent on NAPLAN preparation detracts from time for other curriculum and activities
- II. Outside of schools a wide range of commercial NAPLAN instructional material has evolved, including NAPLAN coaching centres, specialised private NAPLAN tutoring and home study programs. There is little, or no, evidence to verify the effectiveness of these programs.

In their recent review of instructional utility in assessment Young and Kim (2010) equate this concept with formative assessment:

For assessments to be formative—that is, for assessment to be instructionally relevant and the basis for instructional change—teachers need to be able to identify appropriate assessment data (e.g., classroom discourse, observations, tests), use those data to gauge students’ emerging conceptions and individual learning trajectories, and then adjust instruction accordingly. Determining students’ emerging ideas aids the teacher in knowing what parts of previous instruction need additional emphasis, and how to scaffold and tailor subsequent instructional activities.

Initially many teachers, and their professional bodies, wanted to embrace the formative potential of NAPLAN tests, even if questioning their diagnostic credentials. State education departments invested in programs to support and encourage instructional use of NAPLAN data, see, for example, Victoria’s [Using NAPLAN data diagnostically or Western Australia’s NAPLAN Planner resource](#). However, the pressures to *teach to the test* may have surpassed the potential to teach from the test.

Although efforts have been made to make NAPLAN useful to teachers, instructional utility was not present in early aims or design. Issues identified in previous sections of this report contribute to the difficulties teachers face when attempting to make NAPLAN instructionally useful. Previous reviews have identified similar concerns, and none report that teachers are enthusiastic in this regard. Louden’s (2019) recent review of NAPLAN reporting found “*Teachers face barriers in engaging with NAPLAN data and reporting*”. Given that assessment has the potential to be one of the most valuable elements in teachers’ instructional toolbox, NAPLAN presents as a lost opportunity in this regard.

The so-called diagnostic dimension failed to deliver on teachers’ expectations; confusion, criticism and cynicism followed, and the credibility of NAPLAN was called into question (Carter et al., 2018). For teachers, the [timing of the test](#) is the biggest barrier to NAPLAN being instructionally useful. ACARA’s survey of parent views (Colmar-Brunton, 2016) also notes some concerns:

The timing of the tests is also questioned. As they fall in the first half of the year, they do not reflect the entire year of teaching and learning. On the other hand, the tests are not conducted early enough in the year to enable teachers to tailor their teaching to an individual child’s needs. The

perceived usefulness of NAPLAN in this context is also limited by the long time period between sitting the tests and receiving the results.

The impact of the high-stakes pressures of NAPLAN upon pedagogy is unhealthy. High-stakes tests are known to encourage poor pedagogical practice such as drills and rote-learning. There is evidence of this “teaching to the test” dynamic in response to NAPLAN in schools (Dulfer et al., 2012; Thompson, 2012). However, although some positive impacts on teaching and learning have been noted (see Cumming et al.’s review, 2018), few positive trends in outcomes are evident in NAPLAN data, and overall performance has not lifted. Lingard et al., (2016) provide list of unintended consequences noted in research, including: emphasis on performativity, narrowing and compromising of curriculum and pedagogy, teacher and student stress and anxiety, inappropriate interpretation and use of test data, and inequitable treatment of students.

The inclusion of NAPLAN data on the *MySchool* website heightened the stakes of the test and brought a range new uses and unanticipated consequences. These include:

- The use of results for transition from school to school
- The use of results, as part of a student profile, for admission to independent selective and non-selective schools
- The development of a wide range of commercial NAPLAN test-preparation products
- The orientation of coaching schools to NAPLAN-type instruction and test preparation
- The generation of school league tables

So, while teachers find NAPLAN to be less instructionally useful than they would wish, commercial industry was happy to step in with products to support NAPLAN learning. As Jihad Dib, [shadow minister for education in NSW, noted in 2018](#) “*The biggest-selling children’s books are NAPLAN study books, that shouldn’t be the case*”. Little or no research exists examining the pedagogical and developmental soundness of these products; and any instructional benefits they might provide are unverified. These instructional products sit outside the schooling system that NAPLAN should serve, and they are not subject to any form of quality assurance or accountability.

Popham, Berliner and their colleagues (2014) suggest:

Rather than develop tests that first and foremost support accountability decisions and then, as an afterthought, try to figure out how to extract a few morsels of instructionally useful information from these inappropriate tests, we should develop assessments that are designed to help teachers make good instructional decisions. We then need to figure out how (and if) those tests can be used to reasonably support accountability decisions.

Internationally several countries, including Scotland and Singapore, have taken on this challenge. A new Australian assessment system would provide a similar opportunity.

Challenge 6: Fair assessment and marking across schools

Critique of NAPLAN has tended to focus on the purposes and implementation of NAPLAN, rather than the technicalities of the test reliability and validity. Yet these are extremely important foundations for any assessment and needed to meet Australia's key goal of "excellence and equity" in education (Education Council, 2019).

1. Some reports make concerning analyses of NAPLAN's reliability and validity.
2. Annual reports by ACARA show limited disaggregation of results, thus much 'fairness' and the potential for bias goes unexamined and unreported.

There is some suggestion that the high-stakes nature of NAPLAN has led to 'gaming' and in some instances, cheating. Some analyses of NAPLAN provide points of concern regarding reliability and validity; and, in particular, how weaknesses may impact disproportionately upon minority, disadvantaged and low attaining students. Wu (2009, 2015) points out that as only one annual test of 40 questions per subject area, NAPLAN scores contain large margins of error. After examining the reliability and margins of error for all NAPLAN tests she concludes (Wu, 2015):

a NAPLAN test only provides an indicative level of the performance of a student: whether the student is struggling, on track, or performing above average. The NAPLAN tests do not provide fine grading of students by their performance levels because of the large uncertainties associated with the ability measures.

Given the high stakes attached to NAPLAN, this is not an encouraging assessment. Furthermore, the previously discussed issues relating to specific student groups (Indigenous, disadvantaged, LBOTE), mean that the fairness, reliability and validity of tests for some minority groups is likely to be lower than for other students. Parents also expressed this concern and a large proportion reported that NAPLAN "is not as accurate or timely as other school assessments" (Colmar-Brunton, 2016).

In a detailed evaluation of the NAPLAN writing tests Perleman points out a very wide range of issues relating to fairness:

1. A "complete lack of transparency in the development of the NAPLAN essay and grading" (Perelman, 2018)
2. The complete absence of an 'informative writing' genre (NAPLAN examines only persuasive and narrative writing)
3. Marking which uses too many traits. Most writing tests use 3 or 4 but NAPLAN has 10 and a marker has an estimated 270 seconds to make 10 decisions.
4. Writing trait scales that are not balanced. Most tests use similar size scales for different traits, but in NAPLAN the weighting of scales appears arbitrary.
5. Task and marking which puts "too much emphasis on spelling, punctuation, paragraphing and grammar at the expense of higher order writing issues" (ibid.)
6. Tasks which do not define either the audience or the genre for writing – but use marking criteria assessing how both were met.

Perelman's analysis of the writing test is concerning; however, it is unclear if the problems outlined would explain the declines in student writing performance seen over time. It is not difficult to see that many of the flaws he describes would provide major stumbling blocks for students from cultural or linguistic minorities, and for those who are from difficult and socioeconomically disadvantaged backgrounds. Furthermore, low and shifting participation rates, particularly among some minority groups, present a major challenge to valid interpretation of results. Some researchers argue that the results of the NAPLAN are mistakenly well-regarded as a measure of student and school performance despite their low validity; and this is due to complete disregard of participation rates in schools (Thompson et al., 2018).

There is potential, indeed dire need, for more routine and more detailed reporting on NAPLAN to examine fairness, and performance, of various minority groups. Although the national statement of educational goals highlights equity as a key aim, there is currently limited reporting of NAPLAN by equity groups. National annual reporting is limited to gender, and extremely minimal analysis of LBOTE and Indigenous categories. Creagh (2014) points out that the LBOTE category is a catch-all and unhelpful.

Surprisingly, there is no routine disaggregation by socio-economic or school socio-educational status (ICSEA) in reporting of NAPLAN. This is despite the fact that ICSEA data presented for individual schools alongside NAPLAN results on the public MySchool website. A national assessment system that values fairness requires scrutiny, with analyses that look for potential bias that disadvantages particular students. Examining how assessment and attainment relate to socio-economic and socio-educational status is essential if goals of fairness and equity are taken seriously.

Recent independent analysis by Thomson (2021) highlights the extreme educational disadvantage of some groups of students, including Indigenous students, rural and remote students and low SES students. This is also despite markedly lower participation rates among those groups. Similar work, with independent examination of trends by equity groupings, found stark gender and SES differentials among the declines in writing performance (Thomas, 2019).

Furthermore, analyses using NAPLAN data show an increase in the socio-educational gradient explaining student scores between 2010 and 2016 (Bonnor & Shepherd, 2017). This suggests that there is increasing unfairness in terms of education outcomes (Hetherington, 2019; Goss et al., 2016). This is not just an issue of how fair NAPLAN is in assessing students; but points to the failure of NAPLAN to effectively analyse and report on broad educational fairness. One way to redress this problem is to more closely monitor performance across student and school groupings – and use that information to directly inform policy to close the gaps. The experiences of Scotland and Singapore, with their focus on identifying and reducing inequity through assessment and classroom AfL may be relevant. Such approaches are particularly important for Indigenous students, where reporting and analyses are currently inadequate and more sophistication is required for properly informed and effective policy (Cumming et al., 2020).

While some student groups appear substantially disadvantaged by NAPLAN, others may seek unfair advantage. A further issue in relation to fairness, is the reported 'gaming' of the system by some schools seeking to optimise school data and enhance their reputation. Surveys of Australian teachers and parents include reports and comments on most of these aspects

of gaming (Matters, 2018; Cumming et al., 2018). Gaming responses are well-understood as inevitable in high-stakes assessment systems (Nichols & Berliner, 2005). A wide range of gaming responses have been identified, including administrator and teacher cheating; student cheating; the exclusion of low-performance students from testing; and counselling of some students to drop-out or change school. Interestingly teaching to the test and narrowing the curriculum are also considered aspects of 'gaming' (Nichols & Berliner, 2005).

Challenge 7: Informative reporting and proper use of assessment

The reporting of NAPLAN has recently been examined in Louden's NAPLAN Reporting Review (2019) which was commissioned by the Education Council. The review document is lengthy and detailed, but executive summary findings are outlined in Annex 1. where they are listed here against the terms of reference⁴. This summary outlines most, but not all issues, in relation to reporting.

Many of the points highlighted in the Louden review are consistent with the issues outlined throughout this report and echo concerns that NAPLAN is producing negative and unintended consequences; and is not well designed to support teaching and learning. The high-stakes nature and poor timing of NAPLAN hinder, if not defeat, intentions to use NAPLAN for informative purposes.

A key point to be highlighted is that the major barrier to informative reporting of NAPLAN is the untimeliness of the tests and reporting. This has not been identified front and centre in the Louden review, but it is more clearly identified in research and any discussions about NAPLAN with teachers and parents. Teachers identify issues with the timing of NAPLAN tests and reporting; and the consistency between NAPLAN and other assessments (Pierce et al., 2013; Renshaw et al., 2013).

Furthermore, NAPLAN currently has no design elements with specific focus on lifting student learning through the process of testing. Such learning can occur, in addition to other formative assessment aspects, when student engagement with the assessment is strong, and students are given real time feed-back on how they are performing as they complete assessment items. Furthermore, students can also learn directly from the assessment process if they are given prompt and detailed summative feedback after the test. Neither of these is provided by NAPLAN. Other barriers to student learning, previously identified in this report, include NAPLAN's summative rather than formative focus, poor timing, and the lack of attention given to the balance between NAPLAN and teacher-based assessment.

NAPLAN's design as a single annual event that attracts substantial community and media attention that may place undue pressure on students and teachers. This in combination with the publication of data on *MySchool*, which has made it high-stakes, may, in some cases, produce assessment conditions which are more likely to detract from learning. For example, in a Queensland survey of over 7,500 parents, more than 50% reported that the tests had made their child anxious or very anxious and this may have detracted from their child's performance (Matters, 2018, p.5). More than 80% of parents also reported that students get unnecessarily stressed by the experience (ibid. p.10).

While NAPLAN assessments frequently use high-quality stimulus material with potential to drive student learning, these resources are underutilised and there is no timely feedback to students designed to further their learning. NAPLAN reports are delivered to parents at least three months after the tests. [Restrictions on teachers' use of the test papers](#) after they have been completed (ACARA, 2020; ICC, 2016) mean that reflective post-test classroom work to maximise learning is difficult, due to security concerns. Furthermore, there is no opportunity for students to take papers home, discuss them with their families and make deep reflection on their performance. The restriction on test-papers is rationalised by requirements for the development of NAPLAN online from item-banks, however this decision privileges the technical construction of the tests, over the educational purposes and possibilities of assessment.

The recent transition to NAPLAN online for some assessments has enabled a shift to semi-adaptive testing and affords the possibility of effective online feedback, being integrated into the test (Shute, Hansen & Almond, 2008), but to-date this has not been developed. There is also potential to embed national assessment into everyday classroom practice, with teacher controlled responsive testing, closer and timelier alignment with curriculum, and online adaptive testing.

Much of the potential to use assessment to drive learning and improve systems, lies within new assessment technologies and new perspectives on pedagogical integration of teaching, learning and assessment. Online assessment platforms now enable flexible, point-in-time, responsive testing, that can be integrated into day-to-day classroom teaching and learning activities (Pellegrino & Quellmalz, 2010; Beller, 2013; Behizadeh & Lynch, 2017). Such approaches provide a softer touch to testing, with assessment seamlessly integrated into classroom activities providing more authentic classroom assessment conditions and reducing performance pressures upon students. Recent innovations also provide additional focus on learning, with, for example, technology enabling inclusion of immediate and direct feedback to students while they complete large-scale tests. This supports student learning within the assessment process and research has demonstrated that such approaches do not compromise reliability and validity (Shute et al., 2008).

Challenge 8: Ensuring assessment is informed, valued and of optimal utility to teachers

This is, perhaps, where NAPLAN is weakest:

1. There is recognition by teachers that NAPLAN plays a role in monitoring the progress of the system and also in accountability, but research evidence suggests this has not been translated into effective policies able to meet national goals of excellence and equity in education.
2. Because of initial and ongoing confusion on the purpose of NAPLAN much of the teaching profession has felt frustrated and disillusioned by the tests.
3. The full range of issues identified in this report have presented as professional frustrations and challenges for teachers as they seek to minimise negative consequences and maximise teaching and learning opportunities.

Recent calls for reform of NAPLAN, often lead by state and territory ministers for education, have also been matched by calls from teacher professional bodies suggesting “Any replacement testing regime or assessment regime must this time be designed by teachers,” (Mullheron, NSW Teachers’ Federation President, in Baker & Cook, 2019). While national testing has not always required extensive teacher engagement; efforts to strengthen the relationship between national assessments and classroom learning definitely do, and when it comes to Assessment for Learning (AfL) teacher knowledge, skills, beliefs and attitudes are all key to successful implementation (Heitink et al., 2016).

There is some recognition that NAPLAN has contributed to growing acceptance of educational accountability as a necessary professional responsibility among teachers in Australia (Cummings et al., 2018). However, the frustrations that teachers have felt in response to the issues outlined in this report, have dampened any sense of success. Restricted analysis and reporting, with little attention to minority groupings and their trends over time, has also meant that NAPLANs potential for system-level monitoring, and appropriate action for improvement, has been neglected.

From early on surveys of teachers suggested that most see the purpose of NAPLAN as “either a school ranking tool or a policing tool” (Dulfer et al., 2012) that is used as a “stick” rather than a “carrot” (Cummings et al., 2018). Teachers also noted early on that the planning, purpose and design of NAPLAN was not transparently evident. These views have been vindicated with later parliamentary inquiries and national reviews concluding the same (Louden, 2019).

Initially NAPLAN was presented and politically sold with overstated claims as to its potential (Gannon, 2012). In implementation, the timing did not suit the AfL agenda. The so-called “diagnostic” capability was poor due to the broad scale of the tests across the 10 performance bands. More importantly, the tests were high-stakes with discernible and observable negative impacts upon teachers and students (Thompson, 2013; Thompson et al., 2018). Additional burdens were felt by teachers when state and school level initiatives, like [‘Bump it up’ in NSW](#), promoted the achievement of test result targets, for specific target groups, over high quality, autonomous professional practice for the benefit of all students.

There is also some suggestion that the disappointment teachers have felt in relation to the poor formative and instructional utility of NAPLAN, may have had broader negative impact upon teacher engagement with assessment. A recent survey of NSW teachers’ work showed that classroom work associated with external assessment and data collection, analysis and reporting, were among the activities most teachers thought deserved less time and resources (McGrath-Champ et al., 2018). Some qualitative data suggested that classroom assessment is not a priority among teachers as, due to primacy of NAPLAN, teachers reorient their assessment activities for NAPLAN preparation. Many teachers’ comments reflected the view that NAPLAN was an affront to their professional practice. Typical views included:

Then there’s NAPLAN. Principals are now introducing ‘special programs’ designed simply to improve student scores. I don’t blame them, largely arbitrary numbers and the manipulation of them seem to be all that this particular state government is interested in. But study on the detrimental effects of such neoliberal obsessions have been well documented. It does not create or encourage intuitive, interested or motivated learners.

Quite the opposite. In closing, I'm not even sure you could call teaching a 'profession' anymore. Professionals are trusted to use initiative. Their experience is valued. In my lifetime teaching has been dumbed down beyond recognition and older teachers pushed aside and ignored. It is an industry whose proponents are now largely used for the purposes of political scapegoating and as bureaucratic 'gatekeepers' against lawyers and parents.

It is hard to discount the similar and repeated views of so many professional teachers, reflected in countless media articles, research studies and NAPLAN reviews. In June 2019 Queensland Education Minister Grace Grace said she was "constantly" approached by parents, principals and teachers with concerns about the NAPLAN tests. In education circles the weight of evidence and sense of consensus that "we could do better" is unprecedented. It is also clear that "to do better" and design an assessment system which capitalises on potential for learning, there needs to be a much stronger partnership between testing authorities and the teaching profession.

PART 3: INTERNATIONAL PERSPECTIVES

This section includes contributions from three education experts who are deeply embedded in four international education systems: Scotland, Singapore, Ontario and Finland. These systems are often cited as jurisdictions that Australia should look at because of their performance in international comparisons but these education systems are also good examples for their investment in assessment expertise. As you will see in these case studies, many countries have relied on international expert panels to advise on reforms. Each contribution reflects on the jurisdictions education system and particularly their philosophy on national assessment and how their national assessment system has evolved over time.

1 Scotland

By Andy Hargreaves

Educational policy, including testing, is often lumped altogether as UK education policy by international commentators. However, since 1998, and the introduction of devolution for Wales, Northern Ireland, and Scotland, educational policy as it affects schools has been the sole responsibility of each nation within the wider UK.

In 2002, Scottish Government decided to abolish the highly controversial, high-stakes standardized tests that had been in operation since the 1990s, and that still prevail in England, and to eliminate school league tables of results along with them. The reasons for abolishing the tests were consistent with research on the effects of the tests in the UK and elsewhere, especially in terms of narrowing the curriculum, and devoting excessive time among teachers and students to test preparation and practice activities <https://www.theguardian.com/uk/2003/sep/26/scotland.schools>.

National review of education

After a national review of education in Scotland in 2002, a team was set up to create a new national curriculum, known as Curriculum for Excellence, which was first introduced in 2004, though the implementation period was much longer. The team was organized by Learning and Teaching Scotland, a quasi-autonomous body for organizing curriculum, assessment and inspections that has since been renamed and reorganized as Education Scotland. *Curriculum for Excellence* is a national curriculum that is not specified in detailed content like the English national curriculum, for example, but in terms of four essential purposes or capacities, which are, successful learners, confident individuals, responsible citizens and effective contributors.

The four capacities are then broken down into a sizeable number of experiences (Es) and outcomes (Os) through rubrics and levels that are meant to indicate increasing levels of proficiency and mastery in ways that are appropriate for different ages. These Es and Os were set out in a 300-page resource book, for the Early Years and Primary 1 (up to age 5), the end of

Primary 4 (up to age 8), Primary 7 (age 11), Senior 1-3 (Age 12-14), and the Senior Phase (up to age 18). They comprised 1,820 Experiences and Outcomes statements for different areas of subject content and also interdisciplinary learning.

Focus on Assessment for Learning (AfL)

According to University of Glasgow Professor Louise Hayward, shortly before the launch of *Curriculum for Excellence*, Scotland became one of the first countries in the world to have a national framework of what has been called Assessment for Learning. Assessment for Learning (AfL) was a concept developed by Professors Paul Black and Dylan Wiliam in England in 1998. Compared to Assessment of Learning, when the learning is over, as in a grade, test score, or exam result, AfL was meant to be a way to provide teachers with information that would give them feedback on their own teaching and how their students were learning in order to respond effectively to their needs, intervene when they were struggling, and help them make progress over time. AfL encompasses concepts like continuous assessment, formative assessment, and diagnostic assessment, and involves multiple quantitative and qualitative processes such as performance assessments, portfolio assessments and peer assessments alongside traditional diagnostic tools.

The Assessment for Learning framework also encompasses a related concept and process of Assessment as Learning. Assessment as Learning first appeared in publication in 2002, in papers, articles, and eventually a book, by my colleague and fellow co-director of the International Centre for Educational Change at the Ontario Institute for Studies in Education, Lorna Earl. It arose out of a project we conducted together in the mid 1990s on the implementation of a “transition years” policy for students aged 12-14 that addressed the creation of common learning outcomes, interdisciplinary curriculum and alternative forms of assessment. Assessment as Learning identifies assessment as itself being an important domain of learning that young people need to acquire. In Earl’s (2006) own words:

Students, as active critical thinkers make sense of information, relate it to prior knowledge, and use it for new learning. This is the regulatory process in metacognition. It occurs when students personally monitor what they are learning and use the feedback from this monitoring to make adjustments, adaptations and even major changes in what they understand. When teachers focus on assessment as learning, they use classroom assessment as the vehicle for helping students develop, practice and become critical thinkers who are comfortable with reflection and the critical analysis of their own learning.

The principles of assessment for and as learning were integrated into *Curriculum for Excellence* to provide a unitary framework of teaching, learning and assessment. Assessment was meant to be tied to the four capacities and learning within the framework of experiences and outcomes. It was designed to provide evidence of learning and progress, supply information to parents, and contribute to future planning of teaching and learning.

Hard copy and online support (through a National Assessment Resource) provided teachers with descriptors and exemplars of what assessments can and should look like in different areas of learning, specifically in relation to the Es and Os; advice on how to track and monitor students’ progress, and examples of progression in learning. This highly sophisticated and

embedded system was generally welcomed by teachers and teacher associations as it linked assessment closely to learning and to teachers' professional judgment, but by 2015, Education Scotland was sending out signals that the system-wide approach to formative and summative assessment in combination with a profusion of Es and Os, was in danger of being unmanageable and unsustainable.

Further reforms

Large-scale assessments during this period were based on light sampling of literacy and numeracy, much as they are in Finland, and with the National Assessment of Educational Progress in the US, to provide information about progress while avoiding the problems of top-down accountability that existed south of the border. Finland's influence on Scotland's strategy at this time was evident in the award by Minister Mike Russell to Pasi Sahlberg, of the second Robert Owen Award by Scottish Government for service to the country as an inspiring educator. The assessment was known as the Scottish Survey of Literacy and Numeracy (SSLN) that was administered at P4 (age 9), P7 (age 12) and S2 (age 14).

Some significant movements occurred in Scottish Education, including assessment policy, after the election of Nicola Sturgeon as leader of the Scottish Nationalist Party and appointment as First Minister (equivalent to Prime Minister) in November 2014. Sturgeon is the daughter of a dental nurse and electrician, and with her working class roots, has a publicly displayed passion for education and for equity.

After Sturgeon's accession, education took centre stage in Scottish education under the supervision of a new Minister, John Swinney, who was and still also Deputy First Minister – a sign of the high priority attached to educational reform in Scotland. In 2015, opposition politicians began to raise questions about falling standards in Scottish education, Scottish Government commissioned an OECD Review of Curriculum for Excellence, including its assessment strategy, and Sturgeon readied herself to launch a new National Improvement Framework for education in January 2016.

In a submission to a government inquiry into the SSLN in 2018, the Educational Institute of Scotland, the national teacher's union, defended it as a "useful sampling tool" before it "fell foul of the Scottish Government's rash reaction" to political criticism that "sought to capitalise on what was a relatively modest – albeit concerning- dip" in results in 2015. This modest dip, the EIS explained, had occurred in the midst of a dramatic increase in national poverty levels following the 2008 global economic recession and the imposition of austerity measures by the Westminster Government for a decade after that.

Nonetheless, concern about attacking poverty-related attainment problems and increasing educational equity led Scottish government to request a curriculum and assessment review from the OECD in 2015. I was one of two appointed experts within the OECD team of four that conducted its review visit of Scotland in June 2015. The other was Professional Helen Timperley, from the University of Auckland in New Zealand, a leading international assessment expert.

The OECD team completed its report later in 2015. Its conclusions about assessment read as follows (OECD, 2015):

Trust towards teachers and the formative emphasis, with diverse methods to collect information, are very positive aspects of the system. Even in CfE's own terms, however, problems are apparent. School inspection reports express concern that insufficient use is made of assessment information to support children's learning progress and curriculum development. Too many teachers are unclear what should be assessed in relation to the Experiences and Outcomes, which blurs the connection between assessment and improvement. Beyond existing terms, current assessment arrangements do not provide sufficiently robust information, whether for system-level policy-making, or for local authorities, or for individual schools or across CfE domains for learners and their teachers.

Although the principle of integrating formative assessment with curriculum and pedagogy was endorsed, and the argument in favour of high stakes accountability at the national level was rejected on the basis of international evidence, the OECD team noted many problems. Use of moderation among teachers to provide consistency and quality of teacher judgments, although supported by descriptors and examples, was ultimately dependent on professional support at the level of each of the 30 local authorities, and this was highly variable. Evidence collected by Scottish government, by the inspection service, by researchers like Louise Hayward, and by the OECD's own school visits, pointed to insufficient development in teachers' own assessment expertise, not enough use of assessment to support and stimulate children's progress, and the tendency for assessment issues not to take priority in schools in relation to other priorities.

Similarly, at the national level, because the SSLN could not provide sufficient information about strengths and weaknesses at the local authority level, and because it was not integrated with CfE priorities, local authorities ended up designing or adopting their own assessments to fill in the gap, which only hindered national comparability.

The review team proposed lightening the potential assessment load by streamlining the Es and Os. It advised retaining the strong formative focus by building a more robust system of moderation across teachers, schools and local authorities to create consistency of judgment and collective responsibility for improvement. It also proposed revisiting national assessments through a shared approach among local authorities and noted that drafts of a new National Improvement Framework it had seen offered some promise in this regard.

OECD invited Scotland to draw on its rich commitment to CfE, on the high value it placed on teachers' professional judgment, and on developing a robust system of collaboration "to lead the world in developing an integrated assessment and evaluation framework". "It will be essential", the OECD team concluded, "to maintain the dual focus - on the formative function while improving evidence on learner outcomes and progression".

National Improvement Framework

The First Minister and her staff built a National Improvement Framework (NIF) in 2015. The NIF exists alongside and sometimes in tension with CFE as it focuses more precisely on performance in more traditional areas such as literacy and numeracy, and places a high priority on outcomes. It has four priorities:

- Improvement in attainment, particularly in literacy and numeracy

- Closing the attainment gap between the most and least disadvantaged children and young people
- Improvement in children and young people's health and wellbeing
- Improvement in employability skills and sustained, positive school-leaver destinations for all young people

Three of the framework's drivers are related to assessment, improvement and performance.

- **Assessment of Children's Progress:** how we gather information about children's progress and how we use this information to support improving outcomes for all.
- **School Improvement:** every school has a responsibility to evaluate how well it is doing against the National Improvement priorities and other performance measures. These are evaluated by the schools themselves, by the local authority and by HMI inspections
- **Performance Information:** how we gather and analyse information to help us target areas for improvement, show where we have been successful and where we may need to do more.

There was intense dialogue about what large-scale assessment would actually look like within the NIF framework. In pursuing a highly ambitious, aspirational goal of reducing achievement gaps to zero, the First Minister and key members of her staff were initially inclined to reinstate some form of high stakes testing and accountability. Most of the education professionals, meanwhile, alongside the OECD team did not want to see high stakes testing be implemented in ways that would undermine all the progress made within CfE.

A third path was sought and eventually forged which consisted of twin components. The first component, in line with the position of EIS and many elements of the OECD report, was to prioritize teachers' professional judgment but create ways of making it more consistent through structured collaboration. The emphasis here was also on establishing clear levels and descriptors of learning related to CFE goals as a clear basis and framework for the assessments.

The second component, launched in 2017, was to develop an assessment of basic skills that could be administered at various points to all children, but with low stakes rather than high stakes attached. Results would be available nationally to determine changes in national standards, but their prime purpose would be to provide feedback to teachers about their own students and inform their professional judgments. It would be these judgments, developed more formatively - about how well their children were performing at different levels that would provide a basis for intervention and improvement.

Commitment to ongoing evaluation and advice

In 2017, the Scottish government appointed an International Council of Education Advisors (ICEA) to advise the Minister and First Minister on their educational strategy. This rather large group of 10 replicated an earlier experiment with 10 advisers in economics that, the government felt, had been highly successful. The advisers of which I am one are a diverse group of experts, drawn from four continents. The advisers meet twice per year and online in between, engage with various stakeholders, and publish an annual report. This is presented in the public domain,

as is the Minister's transparent response. Assessment issues became part of the ongoing reform agenda that the advisers addressed.

Four items have emerged during the ICEA's deliberations related to assessment.

1. The ICEA advised the Minister to establish Regional Improvement Collaboratives (RICs) among the 30 local authorities where they are grouped in clusters of 5 or 6 to pursue collective responsibility for improvement. Until COVID-19 there were signs of success within these RICs. This has potential for more coherence in formative assessment practices.
2. It urged continuing progress on developing effective moderation of assessments. The point is not to accept professional judgment as it is, but to actively develop professional judgment and make it more consistent by pursuing it collaboratively.
3. ICEA encouraged Scottish Government to see CfE and NIF as two overlapping circles in a Venn diagram, ensuring that one does not eclipse the other.
4. In 2018-2019, after parliamentary opposition forced the government into holding an inquiry into testing in Primary 1, I was invited to present to Scottish Parliament. I informed them that the existing tests had minimal impact, could be taken any time of year, had low stress, did not interfere with the curriculum and did not interrupt play – and had low stakes because their purpose was to inform professional judgment. <https://www.scottishparliament.tv/meeting/education-and-skills-committee-january-30-2019> (At the same time a 2018 EIS review of its members finds that while some teachers find some value in the tests for feedback and professional judgment purposes, in general, they are neither of great benefit nor great harm).

Conclusion

The Scottish experience shows the importance of integrating assessment with learning wherever possible to provide useful feedback to teachers and also develop students' own self-assessment skills.

Teachers place less emphasis on assessment, grading and marking than on the other activities of teaching. So there needs to be a clear structure and culture of moderation (with time and leadership support) to prioritize assessment, and to develop a collective culture of consistency and improvement in assessment practices. This should occur not just within schools but also between them.

Assessment for learning practices should be streamlined wherever possible so they are manageable and sustainable.

We need to continue to ensure that there are reliable data to inform national and local educational leadership. Scotland provides the pathway of having light touch quantitative data of whole cohorts, not samples, to provide national level data and also inform professional judgments that can be a basis for school-based and local information, improvement and intervention.

2 Singapore

By Pak Tee Ng

Singapore is reducing excessive competition and the obsession with examination results

In many ways, Singapore has a robust but traditional examination system. The actual system is more complicated, but a simple way to understand the system is that we have a major national level examination at the end of Primary School (called Primary School Leaving Examination or PSLE), one at the end of Secondary School (GCE 'O' Levels), and one at the end of Junior College (GCE 'A' Levels).

For a long time, parents and the wider society have been preoccupied with examination results. That was because the way to get into good schools, subsequently good paper qualifications and eventually a good life seemed to be determined by examination results. At least that was the general perception or belief. It still is (although this view is no longer as dominant in recent years).

Given that Singapore is a small but densely populated country with many people hoping to get through what they perceived are the common and narrow gates of success, the competition is very keen. Examinations assume an importance bigger than what they ought to do. Therefore, in recent years, Singapore has been making changes to the education system. In particular:

- We are reducing emphasis on examination results
- We are reducing competition among schools and students
- We try to change the mindsets of parents, students, and the wider society regarding education and educational pathways

In some countries, there is a philosophy, based generally on the marketization of education, that school performance on national examinations are to be made public for comparison and that the increased competition among schools will improve education quality. Singapore has had a phase of this, but our current philosophy is quite different. We advocate 'Every School, a Good School' and try to decrease the competitiveness among schools and students. Examinations are still important, but they are not everything! The changes in our examination system must be understood in the light of the changes in the entire education system. The changes in the education system must be understood in the light of the direction of travel of the country as we continue our nation-building journey.

Thinking school learning nation

One of the clearest signs that the traditional model of simply having children get good grades in school is inadequate, is the 'Thinking Schools Learning Nation' (TSLN) vision. Launched in 1997, TSLN is still the umbrella vision for educational initiatives today. To prepare young people so that they could be continually prepared for the future, Goh Chok Tong, who was Prime Minister at that time, explained that TSLN was a vision for a total learning environment, not only in schools but in the whole country (Goh, 1997). In order that the country could thrive in the global marketplace, characterized by increased competition and new technologies, learning had to be

understood as a lifelong process, not just about getting qualifications from schools. Goh (1998) said:

Looking beyond the immediate future, we must focus on lifelong learning and employability for the long term... In fact, the whole country must become a Learning Nation. We must make learning a national culture. We will have to evolve a comprehensive national lifelong learning system that continually retrains our workforce, and encourages every individual to learn all the time as a matter of necessity.

Even so, this cultural shift was not easy because the change was pitted against an entrenched culture. Current Prime Minister Lee Hsien Loong (2013) acknowledged the realities:

Competition is intensifying amongst our students and the focus, unfortunately I think, is too much on examination performance and not enough on learning. It is very hard to fight these forces because parents want the best for their children, and they think the examination results are what make the difference. But I think we need to recalibrate to keep our system open and to focus on things which matter more than exam grades in the long run.

In 2015, as part of Singapore's education strategy, a lifelong learning movement called 'SkillsFuture' was launched to ensure that young people emerge from schools with the skills to deal with complex challenges of the future, and that adult workers have the skills to remain relevant in their various industries (Ng, 2018). SkillsFuture is fundamentally about a change in societal mindset regarding education. Singaporeans are encouraged to learn throughout their lives and learning does not end with schooling. The new direction of the country is to strengthen the intertwining of education and work, so that schools and workplaces are all sites of lifelong learning. Learning is not for examinations. Learning is for life. What are some examples of changes that are relevant to a discussion on examination changes?

Every school is a good school!

In 2012, the MOE introduced the notion that 'Every School, a Good School' to communicate to parents and the wider society that every school in the Singapore education system offered a good education. There was no necessity to compete to enroll in 'elite' schools to receive a good education and achieve success in life. Heng Swee Keat (2012), who was Education Minister at that time, explained:

A 'good school' is one that nurtures engaged learners; enables teachers to be caring educators; and fosters supportive partnerships with parents and the community... each of our schools is good in its own way - as long as we continue to take into account the unique needs and abilities of our students.

So, instead of trying very hard to enrol their children in 'elite' schools, parents should find a school that was suitable for their children. This cultural shift is still work-in-progress. According to Ng (2017, p. 116), the main challenge:

is to redefine what a good school is in the minds of Singaporeans, including educators, parents and the society at large. Good schools

are not just those at the top of ranking tables. A good education does not mean producing the best score in an examination. In a good school, teachers develop their students holistically, not just academically. In a society where every school is a good school, parents focus on their children's overall development instead of just their grades. In the job market, employers hire and promote workers based more on abilities and skills than paper qualifications.

Changes to examinations

In line with 'Every School, a Good School', there have been a few changes to the examination system. Notably, in 2013, Prime Minister Lee Hsien Loong announced that the PSLE scoring system would change. In the past, a student's score in a subject was measured by how he or she has performed relative to his or her peers. In the new scoring system, how well a student did in a subject would be measured using eight Achievement Levels, without reference to the performance of others. The Prime Minister explained that in making this change, students would not need to chase after the last point in a competitive manner. He said (Lee, 2013):

One-point difference in the PSLE scores, 230 versus 231, may make all the difference in your secondary school posting. But at the age of 12, one examination, four papers and you want to measure the child to so many decimal points and say well, this one got one point better than that child? It is a distinction which is meaningless and too fine to make.

With its wider bands, this new system will reduce the 'sharp edges' of examination, and help students feel less pressure as they can focus on their own learning outcomes without comparing with others. In 2018, the MOE announced that it would also be reducing the number of school-based assessments in primary and secondary schools. For example, over the next three years from 2019 to 2021, schools would remove the mid-year exams for Primary Three (P3), Primary Five (P5), Secondary One (S1) and Secondary Three (S3). Also, to discourage excessive peer comparisons, the student report book (officially called the Holistic Development Profile), from 2019 onwards, would no longer present certain academic indicators such as class and level positions of the students (MOE, 28 Sep 2018). These changes reflected the shift in focus from competition to learning.

Increase educational pathways

One of functions of Singapore's system of national examinations (ages 12, 16 and 18) is to serve as a quality control point for learning and a sorting mechanism for young people's future educational pathways. It is neat and tidy, but it also comes with high stakes and high pressure. Therefore, Singapore has been creating more pathways for different types of students to find success in their own ways, and broadening the notion of 'assessment' for the purpose of progression (Ng 2017).

According to Ong Ye Kung, who was Education Minister from October 2015 to July 2020, the end goal of the current education reforms is a multi-path system. Such a system has different types of qualifications and modes of delivery. Certification can be attained at different times and stages of a student's life (Ong, 8 Jun 2018). The official position can be summarized as such (Gov.sg, 2020):

The education system is thus moving towards greater recognition of different aptitudes and talents and encouraging individuals to discover and pursue their strengths and interests, and develop mastery of skills in their chosen fields. In terms of admissions, beyond academic routes, schools and institutions have the flexibility to select and admit students based on their strengths and interests, through the Direct School Admissions (DSA) and Early Admissions Exercises (EAE). The Autonomous Universities (AUs) also practise aptitude-based admissions, by assessing applicants holistically, taking into account a range of factors including their talents and aptitudes.

School rankings

Singapore used to have a league table system that ranked secondary schools according to their academic performance. This system, which started in 1992, was premised on the belief that competition would lead to improvement. Although the system did improve, the competition became too intense subsequently and certain side-effects were counter-productive. Absolute ranking was replaced by broader banding in 2004 to make the system 'gentler'. In 2012, the MOE abolished even the banding of schools. This was a bold and strategic move by the MOE to change the prevailing mindset in the society that schools should be classified according to how their students perform in national examinations (MOE, 12 Sep 2012). Singapore has slaughtered what was widely considered as a sort of sacred cow in the education system.

Changing public perception

In the past, the media focused on news regarding the identities and results of top scorers in national examinations. In order to emphasize that students ought to be recognized for their holistic development and all-round excellence, and not just their examination performances, the MOE has stopped releasing the names of the top PSLE scorers in 2013. Instead, the MOE furnished the media with information about various schools that have performed well in both academic and non-academic areas, and students who have overcome challenging circumstances to do well in their education (MOE, 14 Jan 2013).

All these changes do not mean that Singapore has taken a position that examinations are no longer important. Examination still plays an important role in education. For example, the national year-end examinations will proceed, despite the COVID-19 pandemic, but with safety measures in place. The Director-General of Education explained that "the national examinations are an integral part of teaching and learning, and we also know that the results of the national examinations are important for the progression of the students to the next phase of their educational journey" (Ang, 21 Aug 2020). Minister Ong commented, "I wish I can pass a law and make grades not so important, but that is not for us to decide. It is what society and employers put value on... I think the signal has been sent. Society at large, employers, are beginning to change the way they hire. Many parents, too, are getting the message" (Wong, 28 Mar 2019). So, the change in the education system is not about changes to the examination system alone. Ong said, "This is a period in which the education system is undergoing significant reform, not by one single measure but by a whole package of measures that we have systematically implemented over time. Under this system, we will be able to better prepare our children for the future." (Ng, 28 Jul 2019) So, examination is still important in its own way, but it is not everything!

Conclusion

Singapore's story is one of influencing culture, not just tweaking structures. As Ng wrote (2017, p.124): "In many ways, the education system is a reflection of societal values and expectations. The education system can continue with the current societal norms of competing for schools and focus narrowly on grades and examinations. But, instead, Singapore has decided to act with boldness and resolve to embark on a major transformation of mindsets and influence societal values and expectations."

3 Ontario

By Andy Hargreaves

Ontario's public schools are some of the most admired in the world. For example, in the OECD's 2018 PISA results, Ontario scored equal 5th in the world in reading. This performance is particularly striking given that over a third of young adults in the province are from families where both parents are from another country.

Education Quality and Accountability Office and Literacy and Numeracy Secretariat

Ontario has a complex assessment system. Its most prominent and controversial aspect is the large-scale standardized assessment administered by the Education Quality and Accountability Office, known as the EQAO. The EQAO was established in 1996 as a Crown Agency following the findings of a Royal Commission on Learning in 1995, to respond to growing public demands for accountability, provide independent scrutiny of the effectiveness of the Ontario educational system, and create an assessment system that would apply consistent high standards for learning. It administered its first census-based assessment of Grade 3 in 1997, followed by assessments in literacy and mathematics in Grades 6 and 9. The high school assessment also included the Ontario Secondary School Literacy Test that was first administered in 2002. The passing of the OSSLT was required to achieve a secondary school completion diploma.

By the early 2000s, under Liberal, Conservative, and New Democratic Party (socialist) governments, the EQAO therefore had two common functions of large-scale assessment and examinations – accountability and monitoring, and (in the case of OSSLT), certification. After the advent of a new Liberal Government in 2003, a third function was added that came to be the source of the EQAO tests' problems and controversies. EQAO was now also tasked with having an improvement and intervention function.

Intellectually and strategically, this new purpose arose through the influence of the Premier's special advisor for education, Michael Fullan. Before his appointment as advisor, in 2000, in a paper on "The Return of Large Scale Reform", Fullan reviewed and commented on a number of large-scale reform efforts in New York City, the UK, and the Melbourne region of Australia. Pointing to research by Hill and Crevola (1999) on literacy improvement in a Catholic system in Victoria, Australia, and also on the National Literacy and Numeracy Strategy in the UK (Barber, 2007), Fullan (2000) argued for an interlinked strategy of accountability, targets and testing to drive large-scale improvement. He emphasized the need to:

- "Determine standards and set system-wide and school-specific, year-by-year targets"

- “Focus school support services and available funds on achieving the standards and targets”
- “Put in place accountability and incentive arrangements linked to performance against standards and targets”, and
- “Conduct periodic full-cohort testing to monitor performance against the standards and targets”.

In 2004, Ontario established a Literacy and Numeracy Secretariat. In a quest to improve standards and narrow achievement gaps in literacy and numeracy, the Secretariat developed a four-fold strategy.

First, with an emphasis on the elementary school years, it linked EQAO’s cohort testing in Grades 3 and 6 to province-wide targets for proficiency at level 3 (out of 4 levels). Following the example of the much-criticized UK national testing strategy, a system-wide target was set in which 75% of students had to attain Level 3 proficiency within one election term.

Second, Ontario adopted and adapted the research-and development-based strategy developed by Hill and Crevola to link assessment measures to cycles of improvement and intervention spanning six week periods. Subsequently, in 2006, Fullan authored a book with Hill and Crevola advocating precision of focus and measurement, personalization (or individualization) of micro-attention to student progress, and professional learning to develop teachers’ skills in using this approach (Fullan, Hill & Crevola, 2006). Under this system, as I have witnessed it in my own research over several years with 10 school districts, grade level teams and associated support staff have used proxy measures for EQAO to monitor every individual child’s progress, sometimes on a weekly basis, towards proficiency over 6-week cycles (Hargreaves, 2020). Each child was listed on publicly visible data walls with traffic light indicators. Green meant meeting or exceeding proficiency, red was failing to meet proficiency, and yellow meant at risk of not meeting the standard.

Third, this combination of personalization (individualized monitoring) and measurement-based precision has been used to guide intervention. In 2007, EQAO started to report results that tracked students from Grade 3 to Grade 6. The growing body of system-wide data was also used to provide schools with “statistical neighbours” that were demographically similar but achieving superior results, so they could seek advice and support from them. And the data were used to guide and even drive improvement and intervention at every level from the individual child to the entire province.

In a research report for the Literacy and Numeracy Secretariat, Carol Campbell and Michael Fullan, summarized the approach to intervention and improvement in 8 districts they studied (Campbell & Fullan, 2006). The districts, they said:

Determine standards and set system-wide and school-specific, year-by-year targets.... At district and school levels, there was use of student achievement data, both EQAO assessments and formative assessment data, to evaluate improvement and to identify student groups or curriculum areas requiring particular attention. Working with schools, supervisory officers set clear expectations that principals would discuss with teachers the progress of students and review student data. Within schools,

principals supported teachers in developing practices for classroom monitoring and review, including use of data to assess the effectiveness of their instructional and classroom practices and to modify practices when needed.

Fourth, in an effort to offset the widely reported “perverse consequences” of high stakes testing, the test results were not linked to the kind of punitive consequences that had become widespread in the UK and US – such as terminating principals or closing schools. If schools or districts struggled, they received extensive assistance in the form of advice, resources if needed, and specific coaching help. System-wide, there were also significantly increased resources in terms of training, materials, professional learning, and networking.

To sum up, EQAO was embedded in a four-fold interconnected strategy of target-linked cohort testing, six week-cycles of micro-intervention to raise performance, data-driven intervention at classroom, school, district and provincial levels, and support rather than punishment as a consequence of failure to improve..

Initial achievements and emerging issues

Statistically, in its most intense period of operation up to about 2012, there were clear gains, especially in literacy, and the narrowing of some achievement gaps, especially students with learning disabilities, and second language learners. But almost no independent research was conducted on any other effects of the strategy or related EQAO assessments. Part of the reason for this, perhaps, is that the Ontario Government hired many of the available researchers who had assessment and reform expertise.

The significant exception was research commissioned by the organization representing the 72 Ontario Directors of Education (CODE) carried out by myself in collaboration with Henry Braun between 2008-2011 (Hargreaves & Braun, 2011), and with Dennis Shirley between 2014-2018 (Hargreaves, et al. 2018, Hargreaves, 2020). The research was conducted collaboratively with 10 school districts to do a retrospective analysis of the special education reform strategy and its impact in the first period, and to ascertain the impact and implications of a second wave of reforms during the latter period. Both of these studies unavoidably raise questions about the involvement and impact of EQAO in relation to these strategies.

In the first period, up to 2011, where the government focus was on improving literacy, numeracy and high school graduation rates, our research concluded that the verdict on EQAO was mixed. On the one hand, the Drive to 75 increased proficiency levels to around 70% by 2011. These new levels were up from 54% in 2004 and had increased at steady rates that seemed statistically valid. The tests were also welcomed by schools and system administrators and by special education support teachers. The tests developed a sense of urgency about expectations and equity, they felt, and stimulated data-driven collaborative inquiry to diagnose student difficulties and make interventions in real time.

At the same time, teachers were highly critical of the tests, their unsuitability for high-needs children, and the fact that they were used to get them to focus disproportionate attention on “bubble kids’ falling just below the point of proficiency. This was meant to secure quick gains in the overall scores, though at the expense of students who were having more serious struggles with their learning.

Further reform - achieving excellence

The second period of reform was more expansive and inclusive than improving achievement results in literacy and mathematics. The government set out its new policy in its 2014 report, *Achieving Excellence* (Ontario Ministry of Education, 2014). Areas of excellence like the Arts and STEM were added to literacy and mathematics as priorities. Equity was no longer restricted to narrowing achievement gaps but now encompassed inclusion of diverse and vulnerable groups and their identities, such as indigenous, refugee and LGBTQ students. And student well-being and mental health became a major priority. How did EQAO fare within this broader and bolder educational platform, under the leadership of a new Premier?

In terms of overall results, as the law of diminishing returns set in, EQAO achievement gains became less impressive. Mathematics, which had not received the same attention as literacy, became a new focus of troubling concern. Although pressures to concentrate on “bubble kids” during this second period diminished considerably, support from administrators started to wane and teachers were increasingly vocal about the negative side effects of the testing. These included:

- cultural bias in test items and procedures
- high needs students in extreme poverty or with refugee-related PTSD who couldn't handle the tests
- test preparation activities that took time away from learning
- avoidance of innovation in the years being tested
- unsustainable “spinning cycles” of data-driven improvement given the expanded goals
- anxiety and stress for students and teachers.

Our report concluded that “a 21st century movement in deeper learning and stronger well-being that is embracing a range of innovative practices has been rapidly out-pacing Ontario's 20th century system of large-scale...assessment”. (Hargreaves et al, 2018)

After providing an advanced report of our findings to the Premier and the Minister of Education, on September 6, 2017, they announced an Independent Review of Assessment and Reporting, including the work of EQAO, by the government's six advisors, including Carol Campbell, Michael Fullan and myself. The review – *A Learning Province* - included consideration of submissions of evidence and a range of public engagement meetings across the province involving over 5000 people. (Campbell et al, 2018)

The review recommended that:

large-scale assessment data should not be used for individual student diagnostic or evaluative purposes and students should not be subject to excessive test preparation for a summative system-level snapshot.

It reported the existence of a strong consensus about the negative, indirect side effects of EQAO assessments on students' learning and wellbeing. It advised against “using large scale provincial assessments for student diagnostic purposes, to infer evaluation of educators, and for ranking schools and school boards”.

Among the report's recommendations were:

- Phasing out and ending of EQAO assessments before Grade 6 to avoid threats to well-being among young children
- Allowing flexibility over timing of assessment events for different students
- "Vigilant attention to ensuring curriculum and assessment materials provide linguistically, culturally and geographically relevant items and materials", including indigenous ways of knowing
- Providing effective accommodations and modifications for students with special educational needs including appropriate exemptions
- Mitigating unintended negative consequences especially by sticking to the original purpose of EQAO in providing a summative snapshot rather than also being used as a tool to promote improvements and interventions in particular schools and districts and to rank schools.

The report was presented in March 2018. In April, the Premier accepted all recommendations in principle. One month later, the Liberal Government was not re-elected and the new Progressive Conservative Government disbanded the team of advisors and removed the report from its Ministry website.

In November 2018, EQAO announced minor changes in the test environment. These included broader student access to "headphones, calming white noise or music" and "encouraging the classroom environment to look as it would normally". Other subsequent adjustments include greater accommodations for English Language Learners, and inclusion of Indigenous content.

EQAO testing was suspended in 2020, during the coronavirus pandemic and despite the widespread advocacy for a longer suspension or even complete abolition once students returned to school, on September 23, 2020, Ontario's Progressive Conservative Government announced that it would be introducing an online format for EQAO testing, and invited bids to undertake the work.

Growing Success

The EQAO large-scale test is part of a wider assessment strategy, *Growing Success* (2010), published ten years ago. *Growing Success* was well received by all stakeholders participating in the advisors' assessment review. *Growing Success* is based on seven fundamental principles that support assessment, evaluation and reporting practices that are valid and reliable, and to improvements in student learning. These practices, the report says, should:

- Be fair, transparent, and equitable for all students
- Support all students
- Be carefully planned
- Be communicated clearly
- Be ongoing, varied and administered over a period of time
- Provide ongoing descriptive feedback
- Develop students' self-assessment skills.

Growing Success defines the purposes of provincial assessments as being to “provide data about student achievement to principals, teachers, parents, the public, school district staff, and the government and are used to help educators:

- Identify strengths and areas for improvement in individual students’ learning
- Identify strengths and areas for improvement in the education system
- Develop education policies, allocate resources, and determine the success of those policies and resource allocations
- Identify the need for targeted interventions and supports and provide them where required
- Identify the need for and make decisions about capacity building and specific instructional practices
- Celebrate successes.

Importantly, *Growing Success* warns that:

The results of the provincial assessments should not be used to rank schools or school boards. Rankings tell us nothing about why the scores are high or low. Further, they invite simplistic and misleading comparisons that ignore the particular circumstances affecting achievement in each school and school board. Rankings tend to distract educators and the public from addressing the critical issue of how to improve learning for all students.

The 2018 assessment review noted that *Growing Success* made a clear distinction between the summative purposes of assessment, and diagnostic and summative purposes. The problem, the review continues, is that the EQAO assessments are summative, intended to provide information for use at the level of the overall system (province, school boards, and schools) not for individual students. In practice, EQAO assessments are being perceived and used as diagnostic, formative and summative assessments.

Conclusion

What can be learned from Ontario’s EQAO large-scale assessment? The overall lesson we can take from EQAO is that if a system, government, or electorate wishes to retain large-scale accountability measures of educational achievement, this should be done in a prudent and precise way that does not expand the role of those measures in ways that will incur negative consequences.

In particular, there are eight specific lessons to consider (for further details, see Shirley and Hargreaves, in press):

1. Apply the European Union’s Purpose Limitation Principle that data collected for one purpose should not be used for another. Large-scale assessment can perform valuable accountability and monitoring purposes. It should not be used to drive micro-interventions as if it was a formative or diagnostic assessment with individual students, teachers, and possibly schools.
2. Prevent the use of large-scale assessment data for explicit or implicit ranking of schools or systems (Public, Catholic, Independent) by inhibiting or prohibiting publication of school scores.

3. Create a professional culture and not just a technical system of large-scale assessment in which leaders agree to put an end to unethical testing practices like narrowing the curriculum, teaching to the test, using test scores to attract customers from competing schools, and concentrating undue attention on “bubble kids” to artificially elevate their schools’ scores. This culture can be enshrined in an ethical leadership assessment charter and in the professional standards frameworks for Australian educational leaders.
4. Do not attach testing to time-bound system targets for numerical improvement.
5. In a world where educational goals are expanding and deepening beyond literacy and numeracy as the prime measures of success, ensure that large-scale assessments play a precise and prudent role in overall system-wide assessment strategies.
6. Abandon spinning cycles of short-term, intensive, data-driven improvement based on large-scale assessment instruments or proxies for them.
7. Mitigate risks of perverse consequences by acknowledging, owning up to, and continuously inquiring into the presence and impact of large-scale assessment effects on cultural bias, innovation, student anxiety, and distraction from core purposes.
8. Suspend all large-scale standardized testing before Year 6 on the principle of do no harm, until there is clear and consistent evidence that there are no substantial negative effects on student anxiety levels.

4 Finland

By Pasi Sahlberg

Since the first results of OECD’s Programme for International Student Assessment (PISA) were launched in 2001 Finland has been an admired destination of international education visitors. Before PISA only a very little was known about education in Finland. No wonder, practically all documentation and information about Finnish education was available only in Finnish or in Swedish. What most foreign education experts and journalists realised was that many things in Finnish education system were different compared to much of the rest of the world: Children have less homework, shorter schooldays, and no standardised tests. Rationale behind these odd features of Finnish school education has remained fairly poorly explained in international reviews.

It is difficult to understand specific elements of education systems without understanding the wider context. Just stating that there are no regular census-based standardised tests in Finland without considering the bigger picture of evaluation and assessment policies and practices in education leads to incomplete conclusions. Or, even worse, to claim that Finnish teachers or parents wouldn’t care about or see any value in student assessments, is simply wrong. In order to understand how student assessment system in Finland works and what others could learn from it, we need to look at some of the basic aspects of Finnish education. Three of them are included here: National education evaluation policy, teacher professionalism, and decentralised nature of education governance.

National education evaluation plans

In the 1990s when most other countries in Europe followed the global standardised testing movement that was inspired by whole-system education reforms in England, the Netherlands and United States, Finland chose the other way. The first national education evaluation policy that was endorsed by the education authorities in 1998 and included in the national education legislation (1999) was based on development-oriented evaluation of education systems' effectiveness, efficiency and productivity that assumed that the key role of evaluation of school performance and assessment of student learning is enhancement and positive change, rather than comparison and competition. The national quality assurance system in education that has evolved during the last two decades in Finland and has become an integral part of the Finnish Model of education since then.

The basic elements of the national quality assurance system are the field-specific goals laid down by law, the national core curricula and competence-based qualifications, operating licenses and authorisations to provide education and the regulations for teachers' qualifications. Finnish Education Evaluation Centre (FINEEC) that is an independent expert organisation responsible for conducting and developing evaluation of education operates according to a four-year National Education Evaluation Plan. The purpose of these evaluations is to provide "reliable information for better decision-making and effective steering of the educational policies development on all levels of education". Education Evaluation Plans are formulated by an expert body called the Education Evaluation Council and approved by the Ministry of Education and Culture. First National Education Evaluation Plan was for three-year period of 2009-2011.

For example, the most recent completed Evaluation Plan for 2016-2019 defined four interrelated focus areas with the mutual goals of foresight and impact (Finnish Education Evaluation Centre, 2016):

1. Developing student learning and competencies with evaluation. Evaluations implemented with different enhancement-led methods aim at improving learning results and competence at all educational levels in both official language groups.
2. Functionality and development of the educational system. Evaluation activities that cover all educational levels provide information on the functionality of the entire educational system and policy. The evidence-based evaluation information forms a basis for development work. Evaluations are also targeted at the educational level boundaries and various transition phases.
3. Themes which are central and critical in the society. Evaluations are targeted at societally important and critical themes. Based on an analysis of the changes in the operating environment, significant development targets in education, which are not included in the Evaluation Plan, may be raised for evaluation.
4. Supporting education providers in quality management and in strengthening an operating culture based on enhancement-led evaluation. FINEEC supports education providers and higher education institutions in developing quality management by evaluating their quality systems and producing information on good practices in quality management and development, as well as by spreading the information across different educational levels. Moreover, FINEEC supports schools, educational institutions and higher education institutions in utilising national evaluations and self-evaluations as well as in strengthening the enhancement-led evaluation approach.

In the current Evaluation Plan 2020-2023 the four focus areas are (Finnish Education Evaluation Centre, 2020):

1. Development of learning and skills
2. Strengthening equity of education
3. Enhancement of functionality of the education system
4. Support to continuous development of education.

It is important to stress what are the core purposes of education evaluation in Finland. FINEEC conducts these evaluations primarily for accountability and monitoring the functionality of the education system. More specifically, the evaluations stipulated in the National Evaluation Plans are carried out to serve the following purposes (Finnish Education Evaluation Centre, 2016):

- support the local, regional and national development and decision-making with regard to both the evaluation processes and the evaluation results
- improve the quality of education, the learning of the pupils and students and the work of the teaching staff
- promote the attainment of the goals set for the renewal of the educational system and enhance the monitoring of the development of learning results.

During the current planning period (2020–2023) the FINEEC will organise evaluations in mathematics and mother tongue in three transition points (Year 3, Year 6 and Year 9), in English language in Year 9, and in Finnish as a second language in Year 9. In addition to these assessments, evaluations cover Swedish language, History and Civics, Religious education and Ethics, and Art, Music, Crafts, Domestic science and Physical Education in Year 9. All these evaluations are carried out as sample-based assessments and outsourced to Finland’s research universities and institutions. International student assessment programmes, such as the IEA studies (TIMSS, PIRLS and ICCS) and OECD studies (PISA and TALIS) compliment the findings of the national education evaluations described above.

All evaluations are funded through the state budget to FINEEC. In 2020 the total annual budget that includes staff costs and evaluation projects that cover all forms of education from early childhood education to higher education is less than 3.6 million Euro, or AUD5.8 million. For example, Finnish government invests ten times more in professional development of teachers and improving schools than testing and evaluation of education system.

Teacher professionalism and educational assessment

Since the 1980s the basic academic qualification for teachers in primary and secondary schools has been master’s degree (Sahlberg, 2021). This means about five years of research-based academic study for primary school teachers and typically one additional year of pedagogical studies for secondary school teachers. At the moment, about 97% of primary and secondary school teachers and 99% of school principals in Finland are fully qualified to their current jobs (Finnish National Agency for Education, 2018). Retention rates are very high in Finland and teachers are overall very satisfied in their profession. In brief, professionalism is at very high levels in Finland (OECD, 2019).

Teacher education is offered by Finland's research universities and admission and graduation requirements are set to high overall standards. All teacher education programmes include elements of curriculum, teaching and learning, and educational evaluation. All teachers must be competent in understanding and being able to design student assessments in diagnostic, formative and summative ways.

Because all teachers understand the principles of educational evaluation and are prepared to use different student assessment methods as part of their profession, Finnish teachers have an important role in assessment and evaluation. Evaluation projects mentioned earlier that use sample-based standardised assessments are designed to serve local, regional and system level decision making and quality assurance. Individual student assessments that inform students, parents and schools are the responsibility of teachers and schools. Local governments and individual schools can also use external tests and surveys if that is included in local curricula's assessment and evaluation plans. It is important to understand some peculiar characteristics of student assessment in Finland.

First, students should not be assessed by using numeric or symbolic grades before Year 4 (when students are 10 years old). Until then, reporting to students and parents is descriptive and doesn't compare individual students' performances to one another. Second, all Finnish schools and local municipalities design their own curricula according to the national Core Curriculum for early childhood, primary and secondary education. In these documents general and subject-specific principles of student assessment must be described and their implementation is a requirement for all schools. Teachers shall assess students according to contemporary conceptions of knowledge, learning and teaching, in other words, assessment practices must be based on solid theory and research in educational sciences. Third, assessment practice in schools is balanced between formative and summative assessments that are led by teachers. Formative assessment is an ongoing process during the course of teaching that informs students and parents about progress that students have made. Summative assessments make judgements towards the end of semesters or school year about how well the goals and objectives of learning that are stipulated in school curriculum are achieved. This is the main method of reporting students' progress to parents.

Assessment of student learning is an important aspect of teacher professionalism in Finland. The system leaders trust teachers' collective professional judgment in assessing students' learning at schools. Students are increasingly involved in planning and assessing their own learning in Finnish schools. Interestingly, even when Finnish students' test scores in international assessments, like PISA, have recently declined, the government policies regarding evaluation and assessment of education outcomes have not significantly changed. For example, there is a strong consensus among educators and administrators that census-based national tests would not be the right way to fix the problems that are clearly been identified by various sample-based assessment projects and research during the past decade.

Decentralised education governance

Finland has 310 local governments (municipalities) that are responsible for providing citizens with all public services as described in national legislation, including education. Municipalities own the schools and employ teachers, except a small number of independent schools that have their own system of governance (but are fully funded by central government budget). As

mentioned earlier, each municipality must have a curriculum for their schools that in most cases is designed by the school and approved by local education authorities. Since the municipalities vary from one place to another, so do their curricula.

Municipalities are required by law to monitor and evaluate the performance (efficiency, effectiveness and productivity) of their own education services and take part in external assessments included in National Education Evaluation Plans. Due to autonomy of local municipalities, assessment and evaluation practices also vary from place to place. Local education authorities have also different capacities and resources to design and implement required local assessments. The Matriculation Examination at the end of high school is the only standardized test that all municipalities must organise in centralised way.

Matriculation examination

The only census-based standardised student assessment in Finland is the Matriculation Examination at the end of senior general secondary school (Year 12). It is administrated by an external board appointed by the Ministry of Education and Culture and it has 40 members including academics, high-school teachers, and education specialists from national administration. Exams are prepared and marked by separate subject committees that have altogether over 300 associate members, most of them current or previous schoolteachers. The secretariat of the Matriculation Examination Board has a staff of 22 people. The typical examination fee per student for five exams is about 250 Australian dollars. The entire administration of the examination costs about 15 million Australian dollars annually and it is financed entirely from student-paid examination fees.

Matriculation Examination is organized twice a year and students can complete it in three consecutive semesters of their choice. First, students must take minimum of four individual tests in order to be awarded the Matriculation Examination certificate. An exam assessing students' competency in their mother tongue (Finnish, Swedish, or Sami) is compulsory for everybody. Second, each student chooses three additional tests from the following subject pool: a second domestic language (for example, Swedish), foreign language (most often English), mathematics, and one exam from the humanities and sciences group. Students may also add optional exams as they wish from the following subjects: various foreign languages, history, civics, biology, geography, physics, chemistry, health education, psychology, philosophy, ethics, and religious studies. Students have a maximum of 6 hours to complete each exam, all exams are now fully digital.

Conclusion

Finland's national education evaluation policy is based on enhancement-led methodology in monitoring education system's performance and trust in teachers' expertise in assessing students' learning and progress. Periodical National Educational Evaluation Plans that are prepared by leading education experts and approved by national education authorities provide municipalities and schools with a programmatic framework for local level assessments. Outcomes of education are assessed using sample-based standardised tests, thematic analysis, and in-depth research according to the National Educational Evaluation Plans. International assessment programmes serve as external validation of national evaluations.

There are no plans to change the current national assessment system, for example by moving towards more census-based standardized tests or examinations. Three key conclusions are:

1. National assessment system in Finland is based on enhancement-led evaluation paradigm where measuring student learning outcomes is one piece of the whole
2. National assessment practices are based on 4-year National Education Evaluation Plans that are comprehensive consensus-based programmes linked to current education policies and development needs
3. National assessment of quality and equity of education utilises existing professionalism and assessment capacity of the entire education system by combining the expertise of researchers, policymakers and teachers.

PART 4: CONCLUSIONS

This report has evaluated NAPLAN against a series of challenges that all national assessment systems face. Throughout this exercise there has been an effort to prioritise consideration of how NAPLAN serves students, teachers and parents. A series of issues, and weaknesses, in NAPLAN have been identified.

The issues are not new and have been discussed in many spheres over many years.

Early on the OECD Review of Evaluation and Assessment in Education for Australia identified NAPLAN's unclear aims and poor design for learning as issues, saying (OECD, 2011):

The overall evaluation and assessment framework appears as highly sophisticated and well conceptualised, especially at its top level (national and systemic levels). However, there is a less clear articulation of ways for the national agenda to generate improvements in classroom practice through the assessment and evaluation procedures which are closer to the place of learning.

And suggesting that:

A major step in this direction would be a national reflection about the nature and purpose of evaluation components... and the best approaches for these evaluation components to improve classroom practices.
(Santiago et al., 2011)

This OECD review was completed in 2011, just three years after NAPLAN's implementation, however the points remain salient. Over the last decade shifts in teaching and learning have emerged as unsatisfactory, and in some cases detrimental, by-products of NAPLAN, rather than as strategic policy design to lift educational outcomes.

Some of our findings also echo the concerns about NAPLAN laid out in the most recent review (McGaw et al., 2019), commissioned by ACT, NSW, Vic and Qld governments. As mentioned previously, this review identified NAPLAN's five stated aims, with a distinct orientation to system accountability and school improvement. The interim report bluntly summarised findings as:

- Deficiencies in tests
- Stakes too high
- Curriculum being narrowed
- Results too delayed to be useful.

Other evaluations have identified difficulties in implementation as at the heart of NAPLAN as a wicked problem (Johnston, 2017). Indeed, there is ample research documenting problems with implementation (Dulfer et al., 2012; Klenowski & Wyatt-Smith, 2012; Lingard & Thompson, 2016;

Wu, 2010) and the policy framing of our national assessment system that values, distorts and twists the intents and uses of the tests (Milligan, 2019):

The problem is not in the tests but in the strategy to which the tests belong... NAPLAN is particularly toxic because schools and staff can be [lauded](#), [compared](#), berated and [rated](#) according to their scores, and every child takes home 'the [black dot](#)' report on how they performed on testing day).

These positions have identified problems with NAPLAN in relation to its use as a summative accountability system, and as a system that can be implemented effectively.

This report sought to expand upon this thinking by evaluating NAPLAN against contemporary criteria for national assessment systems and considering how the system has worked for students, parents and teachers. Reflecting on the multitude of issues identified, it is evident that many stem from four key failings:

1. Poor clarity on aims and purposes resulting in use of NAPLAN for a wide range of purposes that it was not designed for, often leading to detrimental consequences
2. A lack of assessment design to support learning and poor balance between formative and summative assessment.
3. Failure of the original intent to make NAPLAN a low stakes assessment
4. Poor initial engagement, later frustration and disillusionment of teachers in relation to NAPLAN due to:
 - Mixed, often contradictory, communication about NAPLAN
 - Frustrations due to implementation difficulties and tests that are not instructionally useful
 - Rhetoric promoting NAPLAN as assessment for student learning, despite obvious barriers to that.
 - Imbalance between NAPLAN and teacher-based assessments

The first point outlines the unhelpful dynamic that has surrounded NAPLAN. From the start the program faced challenges: it was not well integrated, or timed, with other major reforms like the national Australian Curriculum, nor was it designed for many of the purposes to which it was put. As a stand-alone, annual census-based test program NAPLAN takes a singular, one-size-fits-all approach. As we have seen from the international case studies more diverse and comprehensive systems, inclusive of formative and teacher-led assessment, are not only possible but likely to provide better returns. Broader systems enable different assessment sub-systems to meet different purposes.

Related to this, is the second finding which presents as the most tragic loss of opportunity. NAPLAN was not designed for supporting learning. Former ACARA CEO Rob Randall (2014) admitted NAPLAN is not a vehicle for learning, but he argues that NAPLAN is valuable as an evaluation system which is necessary for policy to support learning.

NAPLAN is not in itself a means of improving the quality of education, but it gives us data to have objective discussion about what is working and what is not.

There are two problems with this position. First, focusing on evaluation, misses the opportunities to *also* capitalise on assessment for learning. A system *also* designed for learning has the potential to add great value. Second, this approach, using NAPLAN not to lift learning directly but inform policy-making, has been largely ineffective. Something has been amiss between the NAPLAN reporting and the policy response. Identification of students who do not meet national minimum benchmarks, has labelled them but not resulted in their improvement (Adams et al., 2020). Equity, or minority, groupings have gone unmonitored. Instead, some unintended and negative consequences have emerged.

No recent Australian reforms have satisfactorily addressed the challenges outlined, however, insights can be gained by looking at systems, like Scotland and Finland, where there is greater focus on Assessment for Learning within broader national assessment frameworks; and which are designed to acknowledge and strengthen professional expertise in assessment.

Reforming NAPLAN now, with design for learning, would require a substantial overhaul – and the failings in points 3 and 4 suggests it is not likely to be successful. NAPLAN is high on stakes and low on teacher confidence, thus reinventing it as formative AfL will be a tall order – even with the benefits of online, adaptive delivery. A bolder rethink is required, one which recognises that a composite of several approaches is possible.

One thing is certain, NAPLAN has a swathe of weaknesses clearly identified. The question is: what can be done to redress those weaknesses and create a system that can strengthen, rather than distort practice; and strengthen, rather than weaken Australia’s educational outcomes?

We propose that this can be done by prioritising student, teacher and schools stakeholder needs and, in Part 1 of this report, have outlined a model for a new National Assessment System (NAS). By refocusing on student learning, engaging teachers and supporting schools we will fare better, and make better progress toward our national educational goals of ‘excellence and equity’.

Endnotes

- 1 Although students are referred to by necessity in government assessment reports, the 2019 Measurement Framework for Schooling in Australia (ACARA, 2018) does not refer to teachers at all (with only 1 reference to teaching, 3 to learning, mostly in reference to post-school learning pathways, and one mention of parents). In the most recent NAPLAN report (ACARA, 2019) teachers are mentioned just twice.
- 2 “High-stakes testing in schools is based on the premise that student learning will increase if educators and students are held accountable for achievement. By definition, testing becomes high stakes when the outcomes are used to make decisions about promotion, admissions, graduation, and salaries. High-stakes testing is often associated with public reporting of testing results as a way to bring attention to the assessment results.” From “High-stakes Testing”. In obo in Education. 12 Apr. 2021. <<https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0200.xml>>
- 3 See page 22 and definitions and background below, links are provided for further reading.
Summative “Assessment which takes place at the end of a course of study and provides the final judgement on, or ‘sums up’, the candidate’s performance. The most common form of summative assessment is the end examination.” From “summative assessment.” A Dictionary of Education. Ed. Wallace, Susan: Oxford University Press, , 2009. Oxford Reference. Date Accessed 12 Apr. 2021 <<https://www.oxfordreference.com/view/10.1093/acref/9780199212064.001.0001/acref-9780199212064-e-984>>
Formative Assessment “has roots in curriculum development and evaluation. Cronbach 1963 refers to the idea of using evaluation as a tool for improving curricular programs. Scriven 1967 builds on Cronbach’s work by proposing the term “formative” as a way of clarifying the roles of evaluation. Bloom 1971 applies Scriven’s definition to the process of teaching and learning, by using the term to describe a way of improving student learning. Bloom, et al. 1971 links the idea of formative evaluation to the instructional approach of mastery learning as an instructional process that includes the use of data to improve both teaching and learning. During the 1980s and 1990s, educational researchers continue to expand on the ideas and theories proposed and use of the term “formative evaluation” was replaced by the term “formative assessment.” Sadler 1989 proposes a theory of formative assessment that builds on the definitions previously offered, but it highlights the role of the student in the assessment process and views student self-assessment as critical to improved student learning” from “Formative Assessment”. In obo in Education. 12 Apr. 2021. <<https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0062.xml>>
- 4 Most of this information is a direct quotation from Louden (2019), however, there has been minor paraphrasing for brevity on the four terms of reference and re-ordering of points under those terms. Bolding has been added for emphasis.

REFERENCES

- Acosta, S., Garza, T., Hsu, H. Y., Goodson, P., Padrón, Y., Goltz, H. H., & Johnston, A. (2020). The accountability culture: A systematic review of high-stakes testing and English learners in the United States during No Child Left Behind. *Educational Psychology Review*, 32(2), 327-352.
- Adams, E. K., Hancock, K. J., & Taylor, C. L. (2020). Student achievement against national minimum standards for reading and numeracy in Years 3, 5, 7 and 9: A regression discontinuity analysis. *Australian Journal of Social Issues*.
- Adams, P. (2020) *Moving NAPLAN to the centre of teaching and learning Seminar Series Paper 295*. Centre for Strategic Education. June 2020. Available at: <http://www.cse.edu.au/content/moving-naplan-centre-teaching-and-learning>
- Anderson, J., & Boyle, C. (2015). *Declining NAPLAN participation rates are likely skewing the data*. The Conversation. Available at: <https://theconversation.com/declining-naplan-participation-rates-are-likely-skewing-the-data-42821>
- Ang, H. M. (2020, August 21). *National year-end exams to proceed, students with COVID-19 or in isolation not allowed to sit for exams*. Channel NewsAsia Online. Retrieved from <https://www.channelnewsasia.com/news/singapore/year-end-exams-covid-19-students-safe-distancing-psle-o-a-level-13041330>
- Angelo, D. (2013). *NAPLAN implementation: Implications for classroom learning and teaching, with recommendations for improvement*. *TESOL in Context*, 23(1/2), 53.
- Australian Curriculum Assessment and Reporting Authority. (2008). *National report on schooling in Australia 2017*. Sydney, AU.
- Australian Information Commission (2015). *Marty Ross and Australian Curriculum, Assessment and Reporting Authority [2015] AICmr 75* (1 December 2015). Available at: <http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/AICmr/2015/75.html?stem=0&synonyms=0&query=Marty%20Ross>
- Baird, J., Andrich, D., Hopfenbeck, T. & Stobart, G (2017) Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice*, 24:3, 317-350, DOI: 10.1080/0969594X.2017.1319337
- Baker, J. & Cook, H. (2019) *'It's time': NSW wants NAPLAN replaced with 'genuinely useful' test*. Sydney Morning Herald, June 28, 2019. Available at: <https://www.smh.com.au/education/it-s-time-nsw-wants-naplan-replaced-with-genuinely-useful-test-20190627-p5220m.html>
- Baker, J. (2018) *NAPLAN is being used, abused and must be urgently dumped: Stokes*. Sydney Morning Herald, May 3, 2018. Available at: <https://www.smh.com.au/education/naplan-is-being-used-abused-and-must-be-urgently-dumped-stokes-20180503-p4zd3z.html>
- Barber, M. (2007). *Instruction to deliver: Tony Blair, public services and the challenge of achieving targets*. London: Politico's Publishing.

- Behizadeh, N., & Lynch, T. L. (2017). Righting technologies: How large-scale assessment can foster a more equitable education system. *Berkeley Review of Education*, 7(1).
- Beller, M. (2013). Technologies in large-scale assessments: New directions, challenges, and opportunities. In *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 25-45). Springer, Dordrecht.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles Policy and Practice*, 18(1), 5-25.
- Bonnor, C. & Shepherd, B. (2017). *Losing the game: state of our schools in 2017*. Centre for Policy Development. June 2017. Available at: <https://cpd.org.au/wp-content/uploads/2017/06/FINAL-Losing-the-Game-June-21-with-dedication.pdf>
- Campbell, C., Clinton, J., Fullan, M., Hargreaves, A, James, C., & Longboat, K. D. (2018). *Ontario: A learning province*. Toronto, Canada: Queens Printer.
- Campbell, C. & Fullan, M. (2006). *Unlocking Potential for Learning: Effective district-wide strategies to raise student achievement in literacy and numeracy*. Literacy and Numeracy Secretariat. Government of Ontario. http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/projectreport_full.pdf
- Carter, D., Manuel, J., Dutton, J. (2018). How do secondary school English teachers score NAPLAN? A snapshot of English teachers' views. *Australian Journal of Language and Literacy*, 41(3), 144-154.
- Colbert, P., Wyatt-Smith, C., & Cumming, J. J. (2015). Students at risk and NAPLAN: the collateral damage. In *National Testing in Schools* (pp. 144-156). Routledge.
- Colmar Brunton for ACARA (2018). *NAPLAN research report* (untitled). <https://acaraweb.blob.core.windows.net/acaraweb/docs/default-source/assessment-and-reporting-publications/2018-naplan-online-parent-research.pdf?sfvrsn=2>
- 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/10.1080/17508487.2012.67233>
- Creagh, S. (2014). A critical analysis of problems with the LBOTE category on the NAPLAN test. *The Australian Educational Researcher*, 41(1), 1-23.
- Cumming, J., Goldstein, H., & Hand, K. (2020). Enhanced use of educational accountability data to monitor educational progress of Australian students with focus on Indigenous students. *Educational Assessment, Evaluation and Accountability*, 32(1), 29-51.
- Cumming, J., Jackson, C., Day, C., Maxwell, G., Adie, L., Lingard, B., ... & Heck, E. (2018). *2018 Queensland NAPLAN review: School and system perceptions report and literature review*. Available at: <https://qed.qld.gov.au/programsinitiatives/education/Documents/naplan-2018-school-perceptions-report.pdf>
- Davies, M. (2012). Accessibility to NAPLAN assessments for students with disabilities: a 'fair go'. *Australasian Journal of Special Education*, 36(1), 62-78.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159.
- Donnelly, K., & Wiltshire, K. (2014). *Review of the Australian curriculum: Final report*. Australian Government, Canberra: Department of Education and Training.

- Dowling, A. (2008). *Output measurement in education*. ACER report, December 2008. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=policy_analysis_misc
- Drake, S. M. (2007). *Creating Standards-Based Integrated Curriculum: Aligning Curriculum, Content, Assessment, and Instruction*. Corwin Press, A SAGE Publications Company. 2455 Teller Road, Thousand Oaks, CA 91320.
- 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*. Retrieved from <https://www.whitlam.org/publications/2017/10/17/the-experience-of-education-the-impacts-of-high-stakes-testing-on-school-students-and-their-families>
- Education Review (2019). *NAPLAN tagged stories*. <https://www.educationreview.com.au/tag/naplan/>
- Farvis, J., & Hay, S. (2020). *Undermining teaching: How education consultants view the impact of high-stakes test preparation on teaching*. *Policy Futures in Education*, 1478210320919541.
- Ferrari, J & Bitá, N (2008). *Julia Gillard to follow New York on school ranks*. *The Australian*, August 12, 2008 <https://www.theaustralian.com.au/news/nation/new-york-state-of-mind-on-school-ranks/news-story/0e7829461fcb3d80f5b1f798e7c054a4>
- Finnish Education Evaluation Centre. (2016). *National Plan for Education Evaluation 2016-2019*. Helsinki: FINEEC.
- Finnish Education Evaluation Centre. (2020). *National Plan for Education Evaluation 2020-2023*. Helsinki: FINEEC.
- Finnish National Agency for Education. (2018). *Finnish Teachers and Principals in Figures*. Helsinki: Finnish National Agency for Education.
- Fullan, M. (2000). The return of large-scale educational reform. *Journal of Educational Change*, 1(1), 5–28.
- Fullan, M., Hill, P., & Crévola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin.
- Gannon, S. (2012). *My school re-dux: re-storying schooling with the My School website*. *Discourse: Studies in the Cultural Politics of Education*, iFirst, 1-14.
- Goh C.T. (1998). *Speech by Prime Minister Goh Chok Tong at the May Day Rally at NTUC*
- Goh, C. T. (1997). *Shaping Our Future: Thinking Schools, Learning Nation*. Singapore Government Press Release. Speech by Prime Minister Goh Chok Tong at the Opening of the 7th International Conference on Thinking, 2 June.
- Gorur, R. (2015). The performative politics of NAPLAN and MySchool. In *National Testing in Schools* (pp. 30-43). Routledge
- Goss, P., Sonnemann J., Chisolm, C. and Nelson, L. (2016). *Widening Gaps: What NAPLAN tells us about student progress*, Grattan Institute: Melbourne
- Gov.sg (2020, March 23). *What are the new pathways available in Singapore's education system?* Retrieved from <https://www.gov.sg/article/what-are-the-new-pathways-available-in-singapore-education-system>
- Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, 21, 393–420. Accessed at <https://doi.org/10.1007/s10833-020-09380-5> on December 3, 2020.

- Hargreaves, A. (2008). The fourth way of change: Towards an age of inspiration and sustainability. In A. Hargreaves & M. Fullan (Eds.), *Change wars* (pp. 11-44). Toronto: Solution Tree.
- Hargreaves, A., & Braun, H. (2011). *Leading for all: A research report of the development, design, implementation and impact of Ontario's "Essential or Some, Good for All" initiative*. Toronto, ON: Council of Ontario Directors of Education.
- Hargreaves, A., Shirley, D., Wangia, S., Bacon, C., & D'Angelo, M. (2018). *Leading from the middle: Spreading learning, well-being, and identity across Ontario*. Toronto, Canada: Council of Ontario Directors of Education.
- Harlen, W., Crick, R. D., Broadfoot, P., Daugherty, R., Gardner, J., James, M., & Stobart, G. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. EPPI-Centre, Institute of Education, University of London.
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). *A systematic review of prerequisites for implementing assessment for learning in classroom practice*. *Educational research review*, 17, 50-62.
- Heng, S. K. (2012, September 12). *Keynote Address by Mr Heng Swee Keat, Minister for Education at the Ministry of Education Work Plan Seminar at Ngee Ann Polytechnic Convention Centre*. Retrieved from <http://www.moe.gov.sg/media/speeches/2012/09/12/keynote-address-by-mr-heng-swee-keat-at-wps-2012.php>
- Hetherington, D. (2018). *What price the gap? Education and inequality in Australia*. Sydney: Public Education Foundation. https://publiceducationfoundation.org.au/wp-content/uploads/2018/04/Issues-Paper_What-Price-The-Gap.pdf
- Hill, P. W., & Crévola, C. A. (1999). Key features of a whole-school, design approach to literacy teaching in schools. *Australian Journal of Learning Difficulties*, 4(3), 5–11.
- Howell, A. (2015). *Exploring children's experiences of NAPLAN: beyond the cacophony of adult debate*. PhD Thesis, University of Queensland.
- Johnston, J. (2017). Australian NAPLAN testing: in what ways is this a 'wicked' problem? *Improving schools*, 20(1), 18-34.
- Jones, B. (2008). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*
- Jorgensen, R. 2010. Structured failing: Reshaping a mathematical future for marginalized learners. In *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia*, ed. L. Sparrow, B. Kissane and C. Hurst, 26–35. Fremantle: MERGA. http://www.merga.net.au/documents/MERGA33_Keynote_Jorgensen.pdf
- Klenowski, V. (2015). Questioning the validity of the multiple uses of NAPLAN data. In *National Testing in Schools* (pp. 44-56). Routledge., 23(2), 65-86.
- Lee, H. L. (2013). *Prime Minister Lee Hsien Loong's National Day Rally 2013*. Singapore: Prime Minister's Office. Retrieved from <http://www.pmo.gov.sg/mediacentre/prime-minister-lee-hsien-loongs-national-day-rally-2013-speech-english>
- Lingard, B., & Sellar, S. (2013). 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28(5), 634-656.

- Lingard, B., Thompson, G., & Sellar, S. (2016). National testing from an Australian perspective. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 1–17). London, England: Routledge.
- Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy education in Australia? *Literacy Learning in the Middle Years*, 19(2), 9-19.
- Louden, W. (2019). *NAPLAN Reporting Review*. Prepared for COAG Education Council. Education Services Australia, on behalf of the Education Council.
- Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2018). *The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities*. Language Testing, 0265532218775758.
- Margaret Wu (2009). 'Interpreting NAPLAN results for the layperson', p.2: www.edmeasurement.com.au/_publications/margaret/NAPLAN_for_lay_person.pdf
- Matters, G. (2018). *Queensland NAPLAN Review – Parent Perceptions Report*. Queensland Department of Education, 2018, n.d. Available at: <https://qed.qld.gov.au/programsinitiatives/education/Documents/naplan-2018-parent-perceptions-report.pdf>
- McGaw, B., Loudon, W. & Wyatt-Smith, C. (2019). *NAPLAN Review Interim Report, November 2019*. Reviewed commissioned by the state and territory governments of New South Wales, Australian.
- McNeil, L. M., Coppola, E., Radigan, J., & Vasquez Heilig, J. (2008). *Avoidable losses: High-stakes accountability and the dropout crisis*. 2008, 16. doi:10.14507/epaa.v16n3.2008
- Ministry of Education (MOE) (2012). *MOE Removes Secondary School Banding and Revamps School Awards*. MOE Press Release. Retrieved from <http://www.moe.gov.sg/media/press/2012/09/moe-removes-secondary-school-b.php>
- Ministry of Education (MOE) (2013). *Not Naming the Top Scorers for 2012 PSLE Examinations. Parliamentary Replies*. Retrieved from <https://www.moe.gov.sg/news/parliamentary-replies/not-naming-the-top-scorers-for-2012-psle-examinations>
- Ministry of Education (MOE) (2018). *'Learn For Life' – Preparing Our Students to Excel Beyond Exam Results*. MOE Press Release. Retrieved from <https://www.moe.gov.sg/news/press-releases/-learn-for-life---preparing-our-students-to-excel-beyond-exam-results>
- Mockler, N. (2013). Reporting the 'education revolution': MySchool.edu.au in the print media. *Discourse: Studies in the Cultural Politics of Education*, 34(1), 1-16.
- Ng, C., Wyatt-Smith, C., & Bartlett, B. (2015). *Disadvantaged students' voices on national testing: The submersion of NAPLAN's formative potential*. In *National Testing in Schools* (pp. 152-163). Routledge.
- Ng, J. S. (2019, July 28). *PSLE changes no silver bullet, but part of 'significant reform' to de-emphasise grades: Ong*. Today Online. Retrieved from https://www.todayonline.com/singapore/psle-changes-no-silver-bullet-part-significant-reform-de-emphasise-grades-ong-ye-kung?cid=h3_referral_inarticlelinks_03092019_todayonline
- Ng, P. T. (2017). *Learning from Singapore: The Power of Paradoxes*. New York: Routledge.

- Ng, P. T. (2018). SkillsFuture: The Future of Lifelong Learning in Singapore. In Malone, H. J., Rincón-Gallardo, S., & Kew, K. (Eds.), *Future Directions of Educational Change: Social Justice, Professional Capital, and Systems Change* (pp. 205-221). New York: Routledge.
- Nichols, S. L., & Berliner, D. C. (2005). *The inevitable corruption of indicators and educators through high-stakes testing*.
- O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, 39(1), 4-16. doi:10.1080/03054985.2013.764761
- OECD. (2019). *TALIS 2018 Results*. Volume 1. Paris: OECD.
- OECD. (2015). *Improving schools in Scotland: An OECD perspective*. Paris: Author.
- OECD (2011). *OECD Reviews of Evaluation and Assessment in Education: Australia*. Paris: OECD.
- Ong, Y. K. (2018). *Why Singapore needs a multi-path education system*. Today Online. Retrieved from <https://www.todayonline.com/commentary/why-singapore-needs-multi-path-education-system>
- Ontario Ministry of Education. (2014). *Achieving excellence: A renewed vision for education in Ontario*. Toronto, ON: Queen's Printer for Ontario. Accessed at [www.msdsb.net/images/ADMIN/correspondence/2014/MEDU %20Renewed%20Vision%20for%20Education.pdf](http://www.msdsb.net/images/ADMIN/correspondence/2014/MEDU%20Renewed%20Vision%20for%20Education.pdf) on December 3, 2020.
- Ontario Ministry of Education (2010). *Growing Success: Assessment, evaluation and reporting in Ontario schools*. Toronto, Ontario: Queen's Printer. <http://www.edu.gov.on.ca/eng/policyfunding/growsuccess.pdf>
- Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education*, 43(2), 119-134.
- Perelman, L. (2018). *Towards a new NAPLAN: Testing to the Teaching*. Report commissioned by the New South Wales Teachers' Federation.
- Pierce, R., Chick, H., & Gordon, I. (2013). Teachers' perceptions of the factors influencing their engagement with statistical reports on student achievement data. *Australian Journal of Education*, 57(3), 237-255.
- Popham, W. J. (2000). Big change questions. Should large-scale assessment be used for accountability? –Answer: Depends on the assessment, silly! *Journal of Educational Change*, 1(3), 283-289. doi:10.1023/a:1010054525759
- Popham, W. J., Berliner, D. C., Kingston, N. M., Fuhrman, S. H., Ladd, S. M., Charbonneau, J., & Chatterji, M. (2014). Can today's standardized achievement tests yield instructionally useful data? *Quality Assurance in Education*, 22(4), 300-316.
- Queensland Government (2018) *NAPLAN review*. Department of Education. <https://qed.qld.gov.au/programs-initiatives/education/naplan-2018-review>
- Ragusa, A. T., & Bousfield, K. (2017). 'It's not the test, it's how it's used!' Critical analysis of public response to NAPLAN and MySchool Senate Inquiry. *British Journal of Sociology of Education*, 38(3), 265-286. doi:10.1080/01425692.2015.1073100
- Randall, R. (2014). *NAPLAN tests help us improve vital skills*. The Age, June 12, 2014. Available at: <https://www.theage.com.au/education/naplan-tests-help-us-improve-vital-skills-20140612-zs54n.html>

- Renshaw, P., Baroutsis, A., van Kraayenoord, C., Goos, M., & Dole, S. (2013). *Teachers using classroom data well: Identifying key features of effective practices*. AITSL. Available at: <https://www.aitsl.edu.au/docs/default-source/default-document-library/teachers-using-classroom-data-well.pdf?sfvrsn=0>
- Rose, J., Low-Choy, S., Singh, P., & Vasco, D. (2018). NAPLAN discourses: a systematic review after the first decade. *Discourse: Studies in the Cultural Politics of Education*, 1-16.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11(1), 45-61. doi:10.1007/s10833-008-9098-2
- Sahlberg, P. (2016). Global Educational Reform Movement and its impact on teaching. In Mundy, K., Green, A., Lingard, R., and Verger, A. (Eds.) *The Handbook of Global Policy and Policymaking in Education*. New York: Wiley-Blackwell, pp. 128-144.
- Sahlberg, P. (2021). *Finnish Lessons 3.0: What can the world learn from educational change in Finland*. New York: Teachers College Press.
- Salloum, R. G., Shenkman, E. A., Louviere, J. J., & Chambers, D. A. (2017). Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: a systematic review. *Implementation Science*, 12(1), 140.
- Santiago, P., Donaldson, G., Herman, J., & Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education: Australia*. OECD Publishing (NJ1).
- Shirley, D. & Hargreaves, A. (in press). *Five Paths of Student Engagement: Blazing the trail to learning and success*. Bloomington, IN: Solution Tree
- Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten A hog by weighing It—Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education*, 18(4), 289-316.
- Swaffield, S. (Ed.). (2008). *Unlocking assessment: Understanding for reflection and application*. Routledge.
- Thomas, D. P. (2019). Rapid decline and gender disparities in the NAPLAN writing data. *The Australian Educational Researcher*, 1-20.
- Thompson, G. (2013). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing. *International Education Journal: Comparative Perspectives*, 12(2), 62-84.
- Thompson, G., Adie, L., & Klenowski, V. (2018). Validity and participation: implications for school comparison of Australia's National Assessment Program. *Journal of Education Policy*, 33(6), 759-777. doi:10.1080/02680939.2017.1373407
- Thomson S. (2021). Australia: PISA Australia—Excellence and Equity? In: Crato, N. (ed.) *Improving a Country's Education*. Springer, Cham. https://doi.org/10.1007/978-3-030-59031-4_2
- Tyler, RW (1949). *Basic principles of curriculum and instruction*. Chicago; University of Chicago Press
- von der Embse, N. P., & Witmer, S. E. (2014). High-stakes accountability: Student anxiety and large-scale testing. *Journal of Applied School Psychology*, 30(2), 132-156.
- Whitely, M., Phillimore, J. & Moorin, R. (2020). *The effect of a child's relative age on numeracy and literacy test results: an analysis of NAPLAN in Western Australian government schools in 2017*. Aust. Educ. Res. (2020). <https://doi.org/10.1007/s13384-020-00399-4>

- Wössmann, L., Ludemann, E., Schu ¨tz, G., & West, M. (2007). *School accountability, autonomy and choice, and the level of student achievement: International evidence from PISA 2003*. Education Working Paper No. 13. Retrieved from Paris:
- Wong, P. T. (2019). *Wheels set in motion for attitudes toward academic success to change: Ong Ye Kung*. Today Online. Retrieved from <https://www.todayonline.com/singapore/education-minister-hopes-this-generation-will-change-attitude-towards-academic-success-grades>
- Wu, M. (2010). The inappropriate use of NAPLAN data. *Professional Voice*, 8(1), 21-25.
- Wu, M. (2015). What national testing data can tell us. In *National Testing in Schools* (pp. 18-29). Routledge.
- Wyn, J., Turnbull, M., & Grimshaw, L. (2014). *The experience of education: The impacts of high stakes testing on school students and their families. A qualitative study*. Sydney, NSW: Whitlam Institute.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *education policy analysis archives*, 18, 19.

Annex 1. Louden Review of NAPLAN key findings extracts

1. Perceptions of NAPLAN and My School data, including potential for misinterpretation/misuse

- Australian governments and school systems rely on NAPLAN to make important judgements, but some have reservations about unintended consequences of NAPLAN testing and the publication of NAPLAN data on MySchool.
- NAPLAN provides important but incomplete information on school quality.
- There is no settled view of the purposes or proper uses of NAPLAN and the My School website. Further clarification and communication of these issues would be welcomed.
- Many stakeholders were concerned that publishing school-level NAPLAN data had made the tests “high stakes”, and that any negative consequences flowed from publication of NAPLAN data rather than the collection of data or provision of data to schools and school systems.
- Some, but not all, of the recent Australian league tables have drawn their NAPLAN data from publicly available sources other than My School.
- Parents generally do not place a high priority on NAPLAN results when choosing a school, but many believe that transparency and accountability require publication of these results.
- Concerns about the impact of NAPLAN on wellbeing of teachers and students were reported but estimates of the severity of this impact varied among stakeholders and data sources.
- Some stakeholders advocated for a national accountability system based on sample-based testing, but this is inconsistent with school systems’ and sectors’ current use of population NAPLAN data in their data analytics, school review systems and school board reports.
- There is widespread interest in the development of on-line, formative assessments based on learning progressions.

2. How My School and NAPLAN reporting contribute to understanding of student progress and achievement

- School sector and system data analytics platforms are widely used in understanding student progress and achievement, but schools do not use My School data displays for this purpose.

- Data displays that focus on gain in student achievement were preferred to those that used ICSEA-based calculations to compare statistically similar schools.
- Colour-coding of NAPLAN results was regarded as useful by many focus group participants, but stakeholders had concerns about the use of current ICSEA comparisons as a basis for the similar school calculations that underpin the colour-coding. A technical review of ICSEA would be well regarded
- Measures of student achievement and gain may be sufficient information for public accountability and transparency purposes.

3. How schools use achievement data, including NAPLAN, to inform teaching

- Concerns about the impact of NAPLAN on teaching and learning programs were reported but estimates of the severity of this impact varied among stakeholders and data sources.
- Schools triangulate NAPLAN data with a wide range of other assessments including large-scale nationally normed and standardised tests.
- Schools make limited use of school-level NAPLAN data in communication with families and prefer timelier and more contextual data when discussing individual student achievement.

4. How My School and NAPLAN data are reported to students and parents

- My School and NAPLAN would be strengthened by a public communication program that clarified the purposes and proper uses of the data and the website.
- Students would benefit from age-appropriate explanations of their NAPLAN results.
- The My School website provides more comprehensive and detailed school-level performance data displays than any of the international jurisdictions considered in this Review, but less information than is provided in Australian system and sector data analytics platforms.
- School-level NAPLAN results are widely but not universally available from sources other than My School.
- Parents' views on reporting include some positive aspects but also common negative assessments and many also don't know how to interpret NAPLAN reports.
- Teachers face barriers in engaging with NAPLAN data and reporting

Annex 2. Timeline of online assessment program (source: ACARA)

