2018 Queensland NAPLAN Review

School and System Perceptions Report and Literature Review

Joy Cumming, Christine Jackson
Chantelle Day, Graham Maxwell
Lenore Adie, Bob Lingard
Michelle Haynes, Elizabeth Heck

ACU
INSTITUTE FOR
LEARNING SCIENCES
& TEACHER EDUCATION

Queensland
Government

Contents

Institute for Learning Sciences and Teacher Education 2018

Institute for Learning Sciences and Teacher Education 2018

List of Tables

List of Figures

Institute for Learning Sciences and Teacher Education 2018

EXECUTIVE SUMMARY

Background

This Executive Summary presents main findings from *the 2018 Queensland NAPLAN Review Phase 2: School and System Perceptions—Report and Literature Review* conducted from August 2018 to October 2018. Drawing on views provided by Queensland participants from Government, Catholic and Independent school sectors, this report provides evidence regarding the "optimal positioning of NAPLAN in the future of education in Queensland" and presents findings related to "any changes needed to address issues raised and improve Queensland's education system outcomes" (Department of Education Queensland Government [DoE], 2018). This report also looks at the contribution NAPLAN has made in "enabling Queensland students to reach their full potential and the role NAPLAN plays in school and system improvement" (DoE, 2018). The report builds on Phase 1 of the 2018 Queensland NAPLAN Review, conducted separately to gain parent perspectives with respect to NAPLAN through focus groups in three regions and a state-wide online survey (Matters, 2018).

The authenticity of this report is grounded in the views of the profession and those who work closely in varying capacities as part of the education system. Researchers at the Institute for Learning Sciences and Teacher Education (ILSTE), Australian Catholic University, were engaged to undertake a Literature Review, and to report the professional views of all participants, through online surveys of Queensland teachers, principals, and school students from Years 3 to 10 on the role of NAPLAN, their NAPLAN experiences, and system improvement. In addition, identified key education stakeholders participated in the Review through the platforms of focus groups, interviews and an organisation survey.

The report structure is outlined below.

Chapter 1 Introduction   **Chapter 2 Literature Review**

Chapter 3 Methodology   **Chapter 4 Findings and Discussion**

Chapter 5 Terms of Reference and Key Findings

Major Objectives

The Phase 2 evidence collection was guided by ten issues identified within the fourth Term of Reference. Each of these is provided with the Key Findings.

Methodology

The reviewers employed a mixed-methods approach to the study. Online surveys ensured the maximum number of participants across Queensland were able to contribute their views on NAPLAN, while interviews

and focus groups with key participants were organised across seven regions to ensure all levels of systems and sectors of education had opportunity to participate. Quantitative and qualitative analyses were undertaken of the data. A literature review of both national and international research informed the design of the surveys and provides a further framework for findings and discussion.

Twenty-one nominated key participants took part in interviews and 10 focus groups were conducted across seven regions with 99 participants involved in total. The School Survey was distributed to 70,233 registered teachers, in all school sectors, via emails sent by the Queensland College of Teachers (QCT). Over the available 18-day period, 5,814 completed responses were provided. The Student Survey was sent to all schools via their respective sector representatives with 2,896 responses from students in Years 3 to 10 collected over a 15-day period. The Organisation Survey was completed by 4 participants.

Terms of Reference and Key Findings

ToR 4.1: Value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level

Key Finding 4.1.1: The introduction of NAPLAN in 2008 is seen as a "wake up" call to education in Queensland.

Key Finding 4.1.2: Longitudinal data provide evidence of statistically significant improvement in NAPLAN outcomes for Queensland since 2008.

Key Finding 4.1.3: While Queensland outcomes have continued to progress, they may have plateaued since 2012/2013, as for all states and territories, a common outcome from the introduction of initiatives such as NAPLAN after a period of time.

Key Finding 4.1.4: NAPLAN Writing performance is a concern in Queensland, and nationally, and is an area where further exploration of teaching and assessment format is needed.

Key Finding 4.1.5: NAPLAN has led to acceptance of educational accountability as a necessary professional responsibility.

ToR 4.2: Use, communication and reporting of Queensland NAPLAN data within schools, broader education system and community

Key Finding 4.2.1: Systems and schools engage with NAPLAN data in a variety of ways and to differing extents to monitor student learning and direct teaching. Overall, leaders indicated higher expectations for staff engagement with NAPLAN data than teachers reported in their classroom practices.

Key Finding 4.2.2: Different levels of effective leadership to create collaborative school assessment cultures were evidenced. Key differences related to the extent to which all staff were engaged with senior leaders in examining NAPLAN data and their value for programming and student learning versus selection by senior leaders of the NAPLAN data they considered relevant to teachers. These differences further affected the "buy-in" of all teachers in a school to responsibility for NAPLAN outcomes.

Key Finding 4.2.3: Teachers indicated limited engagement with NAPLAN test data. This was often linked to delays in receiving NAPLAN data for effective use.

Key Finding 4.2.4: Considerable commentary was provided about extensive data collection in schools for triangulation, including NAPLAN data.

Key Finding 4.2.5: Overall, communication about NAPLAN with parents or students was not seen as important by school leaders and teachers. The nature of any discussion regarding NAPLAN was more likely to be about NAPLAN generally, and what it measured, than school or student performance.

Key Finding 4.2.6: While some concerns were expressed regarding the validity of comparisons of school performance on My School, strong concerns were voiced about the inappropriate use of NAPLAN by media for commercial purposes.

ToR 4.3: Expectations, understanding and use of NAPLAN by students, their families, school leaders and systems, and its importance in accountability and monitoring of student outcomes

Key Finding 4.3.1: Phase 2 participants indicated that they had strong understanding of NAPLAN data.

Key Finding 4.3.2: School participants reported little interest in NAPLAN from parents, with exceptions when NAPLAN results were used for entry and selection to a secondary school of the parents' choice.

Key Finding 4.3.3: School staff and student participants indicated little student interest in NAPLAN.

Key Finding 4.3.4: Use of NAPLAN outcomes as performance indicators for middle managers and principals highlights negative accountability uses of NAPLAN data in contrast to effective leadership practices.

ToR 4.4: Factors affecting NAPLAN participation

Key Finding 4.4.1: There is evidence of a decline in participation due to parental concern for their child's wellbeing, and the extent to which they saw NAPLAN as a valuable process. The role of the media in portraying NAPLAN, in terms of school outcomes, teacher professionalism, and reported impact on student wellbeing, was seen as a major influence on parental values.

Key Finding 4.4.2: Parents of students from specific cohorts (EALD, Indigenous, learning needs) held more positive views about NAPLAN and their child's participation.

ToR 4.5: Evidence of the impact of NAPLAN on student and staff wellbeing

Key Finding 4.5.1: There is mixed evidence regarding the extent to which NAPLAN is affecting school personnel wellbeing. School Survey respondents indicated that NAPLAN has a major negative impact on staff wellbeing. However, interview and focus group participants indicated negative impact may reflect overall workload issues or media representations of NAPLAN and effect on school and staff reputation and morale.

Key Finding 4.5.2: The majority of students who participated in Phase 2 of the Review indicated that NAPLAN testing was not having negative impact on their wellbeing. However, there was a considerable proportion of students who reported negative feelings about NAPLAN. This may be affecting engagement with NAPLAN testing. It may also be reflected in the number of parents withdrawing their children from NAPLAN on the basis of anxiety and loss of self-esteem.

ToR 4.6: Effect of NAPLAN on the ability of teachers to teach the full curriculum, school leaders to progress curriculum and program priorities, and schools to deliver on broader educational objectives

Key Finding 4.6.1: School personnel indicated both on the School Survey and through focus groups that attention to NAPLAN and NAPLAN outcomes did affect implementation of the full curriculum. Focus group comments identified impact in terms of reduction of focus on the full Australian Curriculum as well as broader 21st century learning goals.

Key Finding 4.6.2: NAPLAN was seen as representing narrow constructs of literacy and numeracy in terms of the Australian Curriculum constructs of literacy and numeracy, and English and Mathematics.

Key Finding 4.6.3: Interview and focus group participants indicated that the policy discourse with respect to NAPLAN is for schools and teachers to focus on teaching the Australian Curriculum and school assessments. However, evidence with respect to the extent of practice still occurring in some schools, identified from School and Student Surveys, interviews and focus groups, indicates that this has not yet become embedded in practice in all regions and schools.

ToR 4.7: NAPLAN and specific student cohorts, including Aboriginal and/or Torres Strait Islander students

Key Finding 4.7.1: Overall, despite the specific prompts reflecting ToR 4.7, few issues regarding NAPLAN and the achievement of students from specific cohorts, including students who identify as Aboriginal or Torres Strait Islander, students who have English as an Additional Language or Dialect (EALD) and students with disability or special needs, with the exception of wellbeing, were raised by participants.

Key Finding 4.7.2: Educational expectations for students who identify as Aboriginal or Torres Strait Islander stated in policy as national minimum standards were identified as too low; expectations should match those for other students for whom focus is on the upper two bands and As and Bs.

ToR 4.8: Experience of schools and students that participated in NAPLAN Online in 2018

Key Finding 4.8.1: Experiences of NAPLAN Online were both positive and negative. Positive findings related to the increased engagement of most students, accessibility for students with disability, and ease of administration in many schools. Negative findings related to IT infrastructure and internet connectivity affecting not just NAPLAN Online implementation but other school administrative and educational activities for the duration of NAPLAN Online testing.

Key Finding 4.8.2: Further work appears to be necessary in order for teachers and students to develop sufficient computer literacy and keyboarding skills for successful engagement with NAPLAN Online.

Key Finding 4.8.3: Phase 2 participants expressed concern about the form of Online Writing assessment as well as potential impact on student handwriting and cognitive skill development.

ToR 4.9: Impact of NAPLAN on school and system resourcing

Key Finding 4.9.1: Overall, perceptions were that NAPLAN has had positive impact in identifying areas of need at system and school level for further attention, and allocation of resources to schools and curriculum areas.

Key Finding 4.9.2: Some schools may be using financial resources to focus on NAPLAN, for example, through role creation of NAPLAN coordinators, or professional development implicitly focused on NAPLAN literacy and numeracy test score improvement, rather than quality teaching and learning more broadly.

ToR 4.10: Any undesirable consequences for students, teachers, school leaders, schools and the education system

Key Finding 4.10.1: The high stakes accountability of NAPLAN has led to a range of unintended negative consequences and practices for schools, teachers and students in schools. These include allocation of time to NAPLAN test preparation and practice, narrowing of curriculum to focus on NAPLAN elements, and focus on "bubble" students at specific performance levels, in this case, reported to be "upper two bands" or "As and Bs".

Key Finding 4.10.2: Media representations of NAPLAN create a competitive high stakes accountability environment that leads to negative NAPLAN practices.

Key Finding 4.10.3: The extent to which NAPLAN has led to high levels of test-taking in schools using a range of sources may have negative impact on the quality and breadth of teaching and learning over longer cycles.

Overall conclusion

NAPLAN implementation in 2008 created awareness in Queensland of the need to direct attention to student learning in literacy and numeracy. Over time, it has led to improved Queensland performance, in conjunction with increased schooling for children in early learning years. It has served as both a negative and positive driver of education. Current policy emphases in Queensland for schools are strong foci on teaching the Australian Curriculum and school assessment against the curriculum, with NAPLAN seen as one piece of data to inform systems and schools, and parents and the community, about student learning. However, emphasis on NAPLAN as an accountability measure at system and school levels continues to create a negative competitive environment for systems and schools, perpetuating negative educational practices in some schools. Media publication of league tables is seen as creating this environment, distracting schools and teachers from quality teaching and learning practices to suit the needs of learners in the 21st century, recognised in the Australian Curriculum and Melbourne Declaration.

Participants in Phase 2 of the 2018 Queensland NAPLAN Review were relatively comfortable with educational accountability for transparency of educational outcomes and monitoring the health of an education system. They were less confident that NAPLAN in 2018 is still achieving this goal. Ways for improvement of a 21st century-focused accountability system were noted, including the shift of NAPLAN as a census test to a sample test, similar to other National Assessment Program tests. This would necessarily reduce the creation of league tables by media and resultant impacts on practice. System and school personnel noted the need for some indicators of school performance for each school to remain

accountable, but considered other mechanisms may be more suitable. Phase 2 participants identified the need to value and hence gauge educational success in all desired educational outcomes for students. They also appreciated the provision of timely data that assisted in identifying areas of curriculum and individual student learning that needed to be addressed while allowing celebrations of success. Many participants indicated that NAPLAN had served its purpose but it was time for accountability assessment in Australia to evolve.

CHAPTER 1: INTRODUCTION

In 2018, the Queensland Government initiated a review of NAPLAN, cognisant of entering into the eleventh year of implementation throughout Australia. The focus of the NAPLAN Review 2018 is to "ensure Queensland is well placed to participate in any future Education Council commissioned national review of NAPLAN" and for consideration of "the optimal positioning of NAPLAN in the future of education in Queensland, and any changes needed to address issues raised and improve Queensland's education system outcome" (Department of Education Queensland Government [DoE], 2018).

Researchers at the Institute for Learning Sciences and Teacher Education at Australian Catholic University were engaged to undertake Phase 2 of the Review to provide the views, through online surveys, of Queensland teachers, principals and school students from Years 3 to 10 on the role of NAPLAN, their NAPLAN experiences and system improvement. In addition, identified key education stakeholders participated in the Review through focus groups, interviews and an organisational survey.

The Phase 2 evidence collection was guided by issues for the Review to consider, identified in Term of Reference 4:

- the value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level
- how Queensland NAPLAN data is utilised, communicated and reported within schools, the broader education system and the community
- expectations, understanding and use of NAPLAN by students, their families, school leaders and systems, and its importance in accountability and monitoring of student outcomes
- factors affecting NAPLAN participation
- evidence of the impact of NAPLAN on student and staff wellbeing
- the effect of NAPLAN on the ability of teachers to teach the full curriculum, school leaders to progress curriculum and program priorities, and schools to deliver on broader educational objectives
- how NAPLAN affects specific student cohorts, including Aboriginal and/or Torres Strait Islander students
- the differentiated experience of schools and students that participated in NAPLAN Online in 2018
- the impact of NAPLAN on school and system resourcing and
- any undesirable consequences for students, teachers, school leaders, schools and the education system (DoE, 2018).

Stakeholders consulted throughout Phase 2 included: students, teachers, and principals; principal associations; schooling sector representatives; teacher unions and other relevant staff associations; curriculum authorities; and, higher education representatives.

Phase 2 of the Review was comprehensive. Schools, students and organisational representatives from all sectors were included. The core methodological approach taken, detailed in Chapter 3, Methodology, was an invitation to all Queensland registered teachers to complete the online School Survey, with the link embedded in the email. Students were invited to participate in the Student Survey through a survey link and information distributed by email or newsletters by schools. Other key stakeholders were invited directly to participate in interviews or focus groups. Focus groups were undertaken in each of the seven Queensland school regions: Far North Queensland, North Queensland, Central Queensland, Darling Downs South West, Sunshine Coast, South-East, and Metropolitan regions. Three additional focus groups were held with key stakeholders in Brisbane.

This report synthesises the findings that emerged from this comprehensive collection of views and evidence from all participants. The findings are considered and discussed in terms of an international and Australian review of literature on the impact of external testing such as NAPLAN as a system to monitor school and student performance.

It is of relevance to note that the majority of data collection for this project occurred over the period 3 September 2018 to 10 October 2018, with online surveys open during the period 31 August to 17 September 2018. During this period, schools and students received their performance data for NAPLAN 2018. Comparative data on school performance were released by the Queensland Curriculum and Assessment Authority, newspapers in regional towns and cities were publishing "league tables" for schools in their area, and on 16 September, *The Sunday Mail* published a full "league table" for all schools in Queensland (The Sunday Mail, 16 September 2018, pp. 55–58). Several accompanying media articles on Queensland and school NAPLAN data were published during this time. The extent to which the findings presented in this report may have been affected by this timing is not known. However, the findings do reflect the views and perceptions of participants at the time of the data collection.

The structure of the report is as follows:

Chapter 1 Introduction

Chapter 2 Literature Review

Chapter 3 Methodology

Chapter 4 Findings and Discussion

Chapter 5 Terms of Reference and Key Findings

CHAPTER 2: LITERATURE REVIEW

This Review examines NAPLAN in its 11th year of implementation. To examine the perceptions of participants in the second phase of the 2018 Queensland NAPLAN Review, we first situate NAPLAN within its historical context. We then explore: international research literature with respect to accountability assessments and their impact and discuss new international directions in assessment and accountability; and, Australian research literature that has investigated implementation of NAPLAN in school settings. From these, we identify benefits and issues that are used to inform interpretation of our empirical data analyses and findings, and to draw conclusions.

The origins of literacy and numeracy testing in Australia

Literacy and numeracy, and correspondingly English and Mathematics, have been priority learning areas within Australian state and territory education policies and curriculum for several decades. Focus on these areas as common national educational priorities emerged through the development of the first agreement by the Australian federal, state and territory ministers of education on national goals for Australian Education, the *Hobart Declaration of 1989* (Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA], 1989). The primary purpose of the Hobart Declaration was to establish national education goals that would "assist schools and school systems to develop specific objectives and strategies, particularly in the areas of curriculum and assessment" (unpaginated). These would enable all students "to achieve high standards of learning and to develop self-confidence, optimism, high self-esteem, respect for others and achievement of personal excellence". Through the Hobart Declaration, all ministers reached consensus to address development for all students of skills in "English literacy, including skills in listening, speaking, reading and writing" (Aim 6a) and in "numeracy, and other mathematical skills" (Aim 6b).

The Hobart Declaration included the goal for "equality of education opportunities, and to provide for groups with special learning requirements" (Goal 3). Thus, the primary goals were quality education and equity, with learning goals and student wellbeing for all students. The Ministers also agreed to produce an annual National Report on Schooling in Australia from 1990 "marking the beginning of a process of national reporting to the Australian people" to "monitor schools' achievements and their progress towards meeting the agreed national goals". The Hobart Declaration therefore initiated at the national level, in conjunction with learning and equity goals for students, a focus on educational accountability that would "increase public awareness of the performance of our schools as well as make schools more accountable to the Australian people". Work on common curriculum through collaborative development across Australian also began "but [with] no system … bound to use it".

Further national literacy and numeracy policies and agreements were developed in the late 1990s. Policies linked literacy and numeracy learning goals with assessment of students by teachers:

> … as early as possible in the first years of schooling … to ensure that … literacy needs of all students are adequately addressed and to intervene as early as possible to address the needs of those students identified as at risk of not making adequate progress towards the national … literacy goals. (MCEETYA, 1997a)

Clear focus on individual students and students "identified as at risk" was evident in this statement. Additionally, it was proposed that "rigorous State-based assessment procedures" should be undertaken against the benchmark (minimum literacy) standard for Year 3 in reading, writing and spelling from 1998 (MCEETYA, 1997a).

The overall goal was for "every child commencing school from 1998 … [to] achieve a minimum acceptable literacy and numeracy standard within four years (recognising that a very small percentage of students suffer from severe educational disabilities)" (MCEETYA, 1997a). Work commenced by the Curriculum Corporation (CC), in collaboration with school authorities, to develop these benchmark standards for literacy (writing, spelling and reading) and numeracy for Years 3 and 5, and work on standards for Years 7 and 9, for finalisation in 1998 (MCEETYA, 1997b). At this stage, curricula were state, and territory based, not national.

National literacy and numeracy plans were developed in 1998, implementing the benchmarks (Cumming, Kimber, & Wyatt-Smith, 2011, 2012). While the "rigorous" assessment and benchmarks addressed writing, spelling and reading in literacy, the national literacy policy, *Literacy for All: The Challenge for Australian Schools* (Department of Employment, Education, Training and Youth Affairs [DEETYA], 1998), identified "effective" literacy to be "intrinsically purposeful, flexible and dynamic", involving "integration of speaking, listening and critical thinking with reading and writing" (unpaginated), reflecting the range of literacy skills noted in the Hobart Declaration.

The early policy developments leading to implementation of a national literacy and numeracy testing program therefore identified two goals. The earliest national policies focused on individual students acquiring essential skills, with intention for such assessment to be teacher-based. However, this focus contrasted with policy for a national report based on state and territory assessments against stated benchmark standards for school accountability to the community.

The Hobart Declaration was followed by the Adelaide Declaration (MCEETYA, 1999) and the Melbourne Declaration (MCEETYA, 2008). The starting premise for the Adelaide Declaration, similar to the Hobart Declaration, was that national goals provided "broad directions to guide schools and education authorities" to achieve high quality outcomes for all students.

At the student level, the Adelaide Declaration continued commitment to "self-confidence, optimism, high self-esteem, and a commitment to personal excellence" (Goal 1.2), individual student achievement in literacy and numeracy, and equity in outcomes addressing discrimination, disadvantage and opportunity, specifically noting Aboriginal and Torres Strait Islander students.

At the system level, the Adelaide Declaration committed to continuing to

... develop curriculum and related systems of assessment, accreditation and credentialing that promote quality and are nationally recognised and valued ...

increasing public confidence in school education through explicit and defensible standards that guide improvement in students' levels of educational achievement and through which the effectiveness, efficiency and equity of schooling can be measured and evaluated. (unpaginated)

The most recent declaration, the Melbourne Declaration of 2008, sited national education goals in a changing educational and world environment, maintaining aspirations for equity and excellence for all students. Indigenous student outcomes were highlighted as a key priority, as well as outcomes for students from low socioeconomic backgrounds. Literacy and numeracy, in conjunction with other 21st century learning goals, were emphasised. In conjunction with informed citizenship, the second goal of the Melbourne Declaration was for students to be "successful" and "confident", "motivated to reach their full potential" (p. 8), with a positive/resilient "sense of self-worth, self-awareness and personal identity" (p. 9).

The Melbourne Declaration introduced the expectation for world-class assessment practice, reflecting the curriculum and drawing on both "the professional judgement of teachers" and "testing, including national testing", with assessment information on student progress to be used by teachers "to inform their teaching" (p. 14). An overall emphasis in "world-class assessment", as outlined in the Declaration, is on use of assessment evidence to monitor progress, and inform teaching and learning, against curriculum goals and standards.

At the system level, the Melbourne Declaration committed to "strengthened transparency and accountability" for schools and for parents and communities with several statements about the nature of data that should be available (p. 10). While the Declaration identified the role of data to improve teaching and learning within schools, it strengthened narratives regarding individual school performance and comparability.

Schools need reliable, rich data on the performance of their students because they have the primary accountability for improving student outcomes.

Good quality data supports each school to improve outcomes for all of their students. It supports effective diagnosis of student progress and the design of high-quality learning programs. It also informs schools' approaches to provision of programs, school policies, pursuit and allocation of resources, relationships with parents and partnerships with community and business.

Information about the performance of individuals, schools and systems helps parents and families make informed choices and engage with their children's education and the school community.

Parents and families should have access to:

– data on student outcomes

– data that allows them to assess a school's performance overall and in improving student outcomes … (p. 10).

Parents, families and the community should have access to information about the performance of their school compared to schools with similar characteristics. Australian governments will work together to achieve nationally comparable reporting about schools.

In providing information on schooling, governments will ensure that school-based information is published responsibly, so that any public comparisons of schools will be fair, contain accurate and verified data, contextual information and a range of indicators. Governments will not themselves devise simplistic league tables or rankings and privacy will be protected. (p. 11)

Critical throughout the development of national goals of schooling and national literacy and numeracy policies, therefore, have been the combined messages of assisting individual students through assessment, including diagnostic assessment, improving teaching and learning, student wellbeing, equity, and school and system accountability and "transparency" for the community.

Implementation of literacy and numeracy testing in Australia

Literacy and numeracy testing, as an outcome of the Hobart Declaration, commenced with tests prepared and implemented at state and territory level, with a mix of full cohort and sample testing in different jurisdictions. The focus was on the proportion of students achieving the national minimum standards, or benchmarks, in each jurisdiction. However, the nature of tests also varied, with some tests focused on the national minimum standards, while others assessed students across a range of performance levels.

To develop national reports, outcomes from these tests were equated statistically, but not without difficulties. A corollary issue with the original literacy and numeracy testing was the changing sources of the standards to be measured by state and territory testing. While the original work on literacy and numeracy benchmark statements occurred from 1998, the final literacy and numeracy benchmarks for Years 3, 5 and 7 were not published until 2000 (CC, 2000). Given the absence of national curriculum at this time, national statements of learning were further developed and published for English and Mathematics, intended to inform and promote consistency in state and territory curricula, and to form the basis for the next stage Years 3, 5, 7 and 9 benchmarks and testing (CC, 2005a, 2005b).

Given the equating difficulties, the federal, state and territory ministers agreed to improve comparability by implementation of common national tests for Years 3, 5, 7 and 9, as a condition of federal school funding (MCEETYA, 2007), from 2008. The national tests were to move focus from only minimum standards to a continuous progression through 10 bands of achievement across the Year levels, with performance at each level to be judged against six bands (MCEETYA, 2007).

The National Assessment Program—Literacy and Numeracy was implemented from 2008, to become the responsibility, from 2010, of the Australian Curriculum, Assessment and Reporting Authority (ACARA), established in 2008 to develop the Australian national curriculum.

While early NAPLAN tests were aligned with reformulations of the national literacy and numeracy benchmarks, ACARA undertook development of a generic literacy capability, including "'descriptions for the end of Years 2, 4, 6, 8 and 10' … to guide the future development of the National Assessment Program—Literacy and Numeracy" (Cumming et al. 2011, p. 55). It has been aligned to the Australian Curriculum in English and Mathematics since 2016, based on expected learning for year levels prior to the year levels of NAPLAN testing, with additional content from the year of testing and following year to "stretch" students as appropriate (ACARA, 2018b).

In 2015, the National Measurement Framework for Schooling in Australia 2015 (Education Council/ACARA, 2015) identified the key performance measures that were to be the focus of reporting of schooling quality, building on the Melbourne Declaration and national agreements. In addition, to reporting on NAPLAN (and NAP sample assessments), indicators were to include student participation through a range of variables, school completion, and equity, especially for students who identify as Indigenous, have English as an additional language or dialect, are disadvantaged by geographic location or socioeconomic background, or have a disability.

In 2018, NAPLAN is described as testing "the sorts of skills that are essential for every child to progress through school and life, such as reading, writing, spelling and numeracy" (ACARA, 2018a). It provides "benefits from the ground up" and "valuable data to support good teaching and learning, and school improvement" to school systems and governments: "students and parents" can "discuss progress with teachers and compare performance against national peers"; "schools" can "map student progress, identify strengths and weaknesses in teaching programs and set goals"; "teachers" are "help[ed]" to challenge higher performers and identify students needing support" (ACARA, 2018c). While not explicitly addressing student wellbeing, these descriptions clearly focus on individual students and teaching and learning with less focus on the school and system level reporting. An overall focus on addressing disadvantage for identified student groups has been maintained.

The introduction of My School in 2008 provided comparative individual school NAPLAN performance data for the public for the first time.

An ongoing criticism of NAPLAN tests, from their original formulation in the benchmark statements to the current descriptors, is their narrow focus on implementation, given the richness of definitions of literacy and numeracy in place in the Hobart Declaration, until the present descriptors of English and Mathematics in the Australian Curriculum (Cumming et al., 2011, 2012). More recent criticisms have commented on the lack of authenticity of NAPLAN testing of literacy and numeracy (Zammit, 2018) compared with literacy activities and necessary skills, including 21st century learning, for the "real world". These issues are raised in later sections of the review of literature.

Professional expectations for principal and teacher understandings of data

Further to the establishment of ACARA with responsibility for national curriculum and the Australian national testing program including NAPLAN, the Australian Institute for Teachers and School Leaders (AITSL) was established in 2010 to promote teacher quality in collaboration with Australian states and territories. A major role of AITSL has been to establish descriptors of professional standards for Australian

school educators. For teachers, the focus of Standard 5 is "Assess, provide feedback and report on student learning". In addition to their own assessments, teachers are expected to be able "to demonstrate capacity to interpret student assessment data to evaluate student learning and modify teaching practice" (AITSL, 2011, p. 5). While expectations for new graduate teachers are limited in scope, the most proficient teachers, Lead teachers, are expected to be able to "[c]o-ordinate student performance and program evaluation using internal and external student assessment data to improve teaching practice" and to "evaluate school assessment policies and strategies to support colleagues with: using assessment data to diagnose learning needs, complying with curriculum, system and/or school assessment requirements" (p. 17). System accountability and data use are therefore forefront in expectations for teachers' developing professional skills. By contrast, the profiles for principals are of a different nature, with principals expected to promote quality learning by students, but also to have a role in "influencing, developing and delivering on community expectations and government policy", while "contributing to the development of a twenty-first century education system at local, national and international levels" (AITSL, 2015, p. 4).

Therefore, from a national perspective, both teachers and principals have critical roles to play within an accountability assessment framework.

Accountability assessments: An overview of international research

Prefacing Australian accountability developments in 1999, then Minister for Education Dr David Kemp indicated that "[t]he community has a reasonable expectation that the massive public and private investment in school education should lead to appropriate improvements in skill levels and general educational attainment of our young people" (Kemp, 1999). Australia's growing policy agency for accountability assessment and transparency reflects international policy developments that have been occurring over a considerable time (Linn, 2000). As Brill, Grayson, Kuhn and O'Donnell (2018, p. 1) have noted, "conceptualisations of accountability tend to reflect the idea that the mechanism *itself* can be a dynamic agent of positive change".

The U.S. has had the National Assessment of Educational Progress (NAEP) in place since 1969, using sample testing of students across the U.S. in different subject areas each year and providing state comparisons in the areas of Reading, Mathematics, Science and Writing. Criticisms of NAEP have been that while such comparisons are made, the U.S. does not have a national curriculum against which common assessments can be undertaken (Chudowsky & Chudowsky, 2010). Questions therefore have arisen about the validity of the assessments and comparability of outcomes. Nevertheless, NAEP outcomes have been used to examine trends and outcomes in the state-based testing systems introduced in later U.S. educational accountability programs. Most notably, the U.S. *No Child Left Behind* (NCLB) 2002 legislation, later reformed as the *Every Student Succeeds Act*, introduced requirements for U.S. states to undertake annual standardised state Reading and Mathematics tests for all students from Grade 3 to Grade 8 and once in high school, and for each school to chart its Annual Yearly Progress against overall student "proficiency" levels set by states. Such results are published not only for states, but also for school districts and schools. While the original intent of NCLB legislation was to focus on individual student performance and improvement and equity for student educational outcomes, over time the comparison of schools, school districts and states became more significant (Cumming, 2012). Many states have introduced standardised

test outcomes as a requirement for school completion and, further, despite recognition of flaws in processes, to judge teacher competence, including imputation from school-based outcomes for teachers who did not teach students completing the tests (see, e.g., Isensee & Butrymosicz, 2012). Linn (2000) has also noted the issue of target setting from the point of commencement of a new system of accountability testing:

> … gains in the first few years following the introduction of a new testing requirement are generally much larger than those achieved after the program has been in place for several years. This tendency raises questions about the realism of some accountability systems that put in place straight-line improvement targets over extended periods (e.g., 20 years). (p. 7)

U.S. accountability testing is high stakes, affecting school funding and school continuity, teacher employment and pay, and student graduation. The social context in the U.S. for such policy and legislation, reflecting greater racial and economic diversity and, importantly, differentiated opportunity to learn that has long underpinned inequity in educational outcomes, is considerably different from the situation in Australia.

Accountability policy and numerous reforms have also occurred in England. In prefacing a review to be undertaken in 2011, the then Education Secretary commented that:

> We know parents support clear, rigorous and transparent testing at the end of primary school, and the OECD has concluded that external accountability is a key driver of improvement in education and particularly important for the least advantaged. So, we must continue to allow parents to know how their local primary schools are performing. (Department for Education (DfE[UK]), 2010)

This statement, oft-repeated internationally, regarding the significant impact of external accountability reflected an earlier OECD report (2008), stated:

> The strongest impact upon student performance was found in regard to the publication of schools' student achievement data. This was found to have a statistically significant positive impact upon student performance even after accounting for all demographic and socio-economic background characteristics and other school institutional and policy or programme characteristics. Fifteen-year-old students in schools that published this student achievement data scored, on average, *3.5 score points higher* [emphasis added] on the PISA science scale than students in schools that did not publish achievement data, all other things being equal. (OECD, 2008, p. 473)

The statement in the 2008 report arguing international support for publication of school data itself drew on another OECD report (OECD, 2006). A general statement in the Executive summary for the 2006 report noted that, once a range of socioeconomic factors was taken into account, there "remained a significant positive association between schools making their achievement data public and having stronger results" (p. 41). While, examination of the 2006 report did not identify how the value 3.5 points was derived, one

table (5.19d: Accountability policies and student performance in science) indicated a 6.6 positive change in score for accountability policies that involved "school posting achievement data publicly". This change was on a scale with a mean of 500. Inevitably, when sample sizes are large, small differences can yield statistical significance, but not necessarily reflect educational importance. There is also the question to be raised as to the meaning of school provision of data publicly, versus the construction of "league tables", discussed later.

Accountability systems in the UK are again high stakes for schools and teachers through publication of school performance data with ensuing impact on school continuation and funding.

Accountability assessments within Canadian provinces have also been identified as having high-stakes consequences for schools, teachers (effectiveness) and students (graduation) (Koch & DeLuca, 2012). Koch and DeLuca identified the issue of "multiple-use" where results from a single assessment are used for multiple purposes (p. 101), frequently for both system and school level accountability and also to guide specific classroom instruction or student progress. Koch and DeLuca's focus was on a theoretical examination of processes for test validation when multiple uses are in operation and test capacity to address such uses simultaneously.

Unintended consequences

Research has for some time reported unintended consequences, predominantly negative impacts, of national assessment systems when they become high stakes for schools and teachers. Several negative consequences have been documented in international research. A recent UK review of several international jurisdictions reviewed literature that presented evidence of test-based accountability impacts in these countries (Brill, Grayson, Kuhn, & O'Donnell, 2018). In their review, they reiterated the negative general findings that emerge consistently in the literature (detailed below): that high stakes accountability measures of student performance become "privileged" over other areas of curriculum, "teaching to the test"; that teachers may focus on students at "borderlines" neglecting other students; and systems may reform curriculum in response to national performance on international tests. Moreover, they noted that "[p]upils may become less engaged learners when undue emphasis is placed upon performance of some groups at the expense of others" (p. ii). However, they also found a "paucity" of evidence about "impact of accountability on the curriculum, standards, and teacher and pupil engagement", and "little robust evidence about accountability on teacher workload, and teacher and pupil wellbeing" (p. ii).

Over two decades, overall, international research attributes a number of negative outcomes to accountability programs and testing, especially as such programs become "high-stakes":

- schools resort to "game-playing" to improve assessment outcomes, even removing students from testing (Heilig & Darling-Hammond, 2008)
- teachers concentrate on students near key borders or benchmarks to get them over the line, leaving behind students most at risk (Bew, 2011; Jennings & Dorn, 2008)
- teachers narrow the curriculum, prioritising test content at the expense of the full curriculum, overemphasising decontextualised skills (DfE(UK), 2010; Harlen, 2005; Kramer-Dahl, 2008; Hursh,

2005, 2008; Spielman, 2017; Stobart, 2008; Stobart & Eggen, 2012) especially in grade levels being tested (Stecher & Barron, 2001) —

> too many schools feel they must drill children for tests and spend too much time on test preparation in Year 6 at the expense of productive teaching and learning (Bew, 2011, p. 7) and overpractise item types on tests with further narrowing of the curriculum (Harlen, 2005; Shepard, 2003), especially, when question formats are predominantly multiple choice and over-rely on simple, highly structured problems that tap fact retrieval and the use of algorithmic solution procedures (Timmis, Broadfoot, Sutherland, & Oldfield (2016) as cited in Pellegrino & Quellmalz, 2010, p. 461).

- teachers make little use of results to assist student learning (Harlen, 2005)
- tests promote passive learning (see Brill et al., 2018)
- testing creates pressure and stress on teachers, test anxiety and/or test aversion for students (see Brill et al., 2018), impacting on student wellbeing including high achieving students (see Brill et al., 2018), and for parents and students, including use of private tutoring and "cram" schools (see Brill et al., 2018; Kwon, Lee, & Shin, 2017)
- testing creates pressure on schools through ranking
- cheating (Amrein-Beardsley, Berliner, & Rideau, 2010).

In conjunction with overall reported concerns about the negative impacts of test-based accountability with limited evidence of improvement in student learning, criticisms have been raised about the focus of such accountability tests due to their focus on "outcomes" rather than "processes of learning" (Baird et al., 2014, p. 6). Such focus on outcomes has also led to use of "data walls" as visual displays of data on performance and progress, for schools, classes and for students (Renshaw, Baroutsis, van Kraayenoord, Goos, & Dole, 2013). Opinions on the value of data walls are divided, despite, or perhaps due to, the ease with which data can be tracked for accountability purposes (Wyatt-Smith, Adie, & Harris, 2018; Wyatt-Smith, Harris, & Adie, 2018). Reflecting the concerns of Baird et al. (2014), the issue is the extent to which data walls present data that are decontextualised from contexts such as intended curriculum goals, student characteristics, how data can be used to inform pedagogy and represent learning and connections with other data. A systematic review of data walls and evidence of learning was undertaken by Wyatt-Smith, Harris, and Adie (2018), as well as consultation with international experts in assessment and decision-making. Twenty-one research articles providing empirical data on data wall use and impact on student learning were identified. Overall, at present limited research evidence is available as to how data walls impacted on student learning and achievement, whether used by school staff or students for self-monitoring. In one study that indicated positive learning outcomes in conjunction with data walls, the use of data walls was part of a much larger context for using data to improve student learning. Other studies report teacher narratives about data wall use and impact. Wyatt-Smith, Harris, and Adie's overall consideration was that data walls were still experimental and that use and impact should be monitored.

The prevalence of such data wall use is indicative of the extent to which schools, teachers and, to an extent, students are focused on achievement outcomes represented by "hard" data, perceived as objective and reliable. Accountability and targets for improvement can encourage a culture of performativity, precipitating a stress on improved test scores over other more educative foci, and ensuring principals and

teachers became extremely responsive to numerical calculations of the outcomes of their work (Ball, 2003). The work of schools and teachers is framed more and more by data and an increased acknowledgement of it. This can narrow the work of schools and teachers, strengthening a culture of performativity and compliance with expectations that student learning, growth and quality education are measured by external and quantitative measures of achievement. This weakens teacher professional knowledge and identities, and respect for teachers' own professional assessment skills and judgements.

Focus of much accountability assessment on such outcomes rather than learning processes is in part attributed to their development by psychometricians described as being "concerned only with the status of the individual in terms of a trait or construct, and not directly with how the person achieved that status" (Ball, 2003, p. 51), making no or limited contribution to learning theory research and development. Standardised tests are developed by psychometricians at some distance from classrooms, curriculum and teachers' professional knowledges and practices. Ball (2003) has argued in the English context that control over the "field of judgement" has passed from teachers to the psychometricians with this mode of test-based accountability. This means technical considerations become more important, or at least as important, as educational considerations in test construction (Gorur, 2016). An ongoing concern remains the validity of such tests in curriculum-focused school systems.

Evidence that accountability assessment has led to learning improvements

Research has noted the educational intention of national tests such as NAPLAN to "focus instruction and learning" on important curriculum content and skills; "define standards and expectations for student achievement"; and "provide teachers and schools with information about student achievement", especially for students needing additional attention (Madaus, Russell, & Higgins, 2009, p. 2). While research on negative impacts of test-based accountability can be criticised as both limited and tending to be based on small qualitative research studies, international research on the effective use of data, including accountability assessment data for teaching and student learning improvement, is also limited and of a similar nature.

Overall, there is little evidence in international research of a positive impact of accountability testing in improving student learning, or that such testing, in conjunction with public reporting, has been successful as a driver in raising student achievement with other than modest outcomes. NAEP analyses of cross-sectional data identify that from 1973 to 2008, a period of 35 years, Reading scores for all U.S. students participating in the NAEP sample testing improved 13 points (on a 500-point scale) for 9-year-old students, 8 points at age 13, and at age 17 had no observed improvement (NAEP, n.d.). While improvement was greater for Black and Hispanic students than White students, Heritage (2014) has noted that this may be at the cost of improved quality of learning experience for these students, as discussed below. Similarly, analyses of outcomes for NCLB over time have shown only "modest" outcomes, "limited in both size and applicability", and in some cases negative (Hout & Elliott, 2011, p. 82).

Test-based accountability, equity and social justice

One significant effect of the rise in policy significance within nations of international large scale assessments such as the OECD's PISA, the IEA's TIMSS and PIRLS, complementary national testing such as NAPLAN in Australia, and their use as test-based accountability, has been the way in which the social justice

goals of schooling, which underpin the work of most schooling systems, including those in Australia, have been rearticulated as equity as measured on these complementary international and national tests (Lingard, Sellar, & Savage, 2014). Importantly, the analysis of PISA performance data on national systems gives emphasis to both equity and quality. Both are defined in terms of numbers and test performance. The strength, though, of the OECD's PISA is that it demonstrates quite clearly that quality and equity go together in high performing PISA schooling systems, rather than sit in tension with each other. Condron (2011) also demonstrates that equity and quality are compatible goals for schooling systems in affluent societies. On PISA, quality refers to the comparative performance of systems on the three tests of Reading Literacy, Mathematical Literacy and Scientific Literacy, emphasising equity as high scores and low standard deviations on each of the tests. Equity is also defined on PISA as the strength of the correlation between students' socioeconomic background and performance on the test. There is an additional numerical rearticulation of equity on PISA tests, namely, the percentage of "resilient students" in each system, defined by the OECD as the percentage of students in the bottom quartile of socioeconomic background who achieve in the top two categories of performance on PISA. We see in this OECD PISA analysis the numerical rearticulation of what social justice is, and its discursive reframing as equity (see Lingard, Sellar, & Savage, 2014, pp. 722–724).

NAPLAN in Australia in a complementary way has also rearticulated social justice as equity and in relation to analyses of performance on NAPLAN. In relation to NAPLAN, equity is rearticulated in a number of metricised ways: one definition is the number of students in a school or school systems reaching the national minimum standards on each element of the literacy test and also on the numeracy tests. A national target has been set at 80% of students reaching these minimum standards at each year level of the test, implying ongoing inequities through schooling. Specific targets for achievement outcomes by Australian Indigenous students are not set in absolute terms but in terms of "halving the gap" in outcomes between Indigenous and non-Indigenous students, that is, in relative terms (Commonwealth of Australia, Department of the Prime Minister and Cabinet [CADPMC], 2018). Equity on NAPLAN is also defined in relation to a school's comparative performance against approximately 60 statistically similar schools. Similar schools are constituted through the Index of Community Socio-Educational Advantage (ICSEA) (incorporating parent education and occupation data and proportion of Indigenous students in a school) for comparisons on My School. ACARA explains that ICSEA consists of a "combination of variables that have the strongest association with student performance on the National Assessment Program—Literacy and Numeracy (NAPLAN) tests" (ACARA, 2015a, p. 1). The rationale for this Index to create Like School measures is one grounded in equity concerns that a students' socio-educational background should not "determine" or "restrict" their schooling and learning opportunities. It needs to be noted that the 60 similar schools' measure does not provide a measure of equity *per se*, but rather provides a comparative measure of performance. This measure, it is argued by ACARA, controls for the different contexts of schools and thus ensures that a school's performance against that of like schools is deemed to be a result of in-school factors, principal leadership, teacher pedagogies and so on. This measure then effectively "responsibilises" schools and their work, denying the impact of structural inequality surrounding schools and thus ensuring a "fatalism" toward such inequality (Power & Frandji, 2010).

Therefore, in the testing work of the OECD with PISA and in NAPLAN in Australia, equity has been redefined in reductive and numerical ways that bracket out structural inequality and which encourage schools to focus on improving test performance. There may be a need to re-tether necessary data concerning social justice in schooling systems with a re-conceptualisation of what social justice in schooling ought to be in today's globalised world (Lingard, Sellar, & Savage, 2014). This would result in the concept driving data collection, rather than data collection redefining the concept as at present. At the same time, this work would need to acknowledge that, "[r]efusing to deal with numbers rarely serves the interests of the least well-off" (Piketty, 2014, p. 577).

Effective use of national testing data to improve learning

Renshaw et al. (2013) provide a brief overview of literature related to interpretation and use of data, including NAPLAN type assessment data, to improve teaching and student learning, noting the considerable amount of data available within schools. They cite literature that identified several barriers to effective use of data, including "cultural"—when teachers prefer to use their own experiences to judge student progress, "technical"—when data are not available or are not appropriate, and "political"—when overpoliticising of data leads to resistance and mistrust (p. 29). More specific issues in school and teacher use of data relate to their knowledge and understanding of such data, and capacity to interpret it in context and integrate different sources of data or evidence. As Renshaw et al. (2013) noted in their own study, Queensland teachers were involved in an avalanche of data precipitated by NAPLAN, interpreted narrowly outside classroom assessment, with further concerns that such data were not used effectively to promote teaching and learning.

Overall, research has identified that effective evidence-informed practices for classroom use of external test data to improve learning, include:

- availability of diagnostic supports (Darling-Hammond, 2003)
- professional development that assists teachers to support students (Wyatt-Smith, Bridges, Hedemann, & Neville, 2008)
- a teacher or teachers work as "data gurus" within schools (Boudett, City & Murnane, 2006; Cromey & Hanson, 2000)
- teachers use information to work with students and chart their progress (Black & Wiliam, 1998; Holmes-Smith, 2005)
- assistance is tailored to individual needs through an inductive approach rather than through a deductive response using ***a priori*** developed programs (Black & Wiliam, 1998).

Ikemoto and Marsh (2007) identified seven factors that have been shown to be influential for the effective use of data in schools:

- accessibility and timeliness of the data
- perceived validity of the data
- professional capacity and support
- tools for data analysis
- external support and expertise

- time to analyse and interpret data and decide what action to take
- leadership and culture.

The last of these is recognised as essential to achieving the other factors, and hence of primary importance (Cumming, Maxwell, & Wyatt-Smith, 2016; Maxwell, in preparation). Critical for data use is establishing a collaborative culture of inquiry (Earl & Katz, 2002) and support (Anderson Leithwood, & Strauss, 2010). As Maxwell (in preparation) has noted, elaborating work by O'Day (2002), Sutherland (2004) and Wahlstrom et al. (2010):

> For deep and long-term improvement in educational outcomes, the external imposition of policy directives and accountability threats does not work; a more professional approach is needed to deal with the complexities of schools and teaching, one that allows for flexibility in exercising professional judgment and applying professional expertise, while remaining accountable for actions taken. External mandates on schools can create an impetus to attend to data on educational outcomes, but unless this is complemented by school practices that value the use of data for improving student learning, these mandates are unlikely to have sustained effects. The extrinsic motivation (externally controlled) created by any accountability regime needs to be complemented by an intrinsic motivation (self-determined), with personal commitment to the principle of data-informed educational improvement.

New international directions in accountability assessments

As well as providing a literature review, Brill et al. (2018) examined six case studies of accountability. These included England, Key stage assessments (and Wales); Australia, NAPLAN; Japan, national assessment at the end of primary school but not for accountability, school self-evaluations and school external evaluations including inspection; New Zealand, the national Monitoring Study of Student Achievement, a sample assessment that does not report on students, teachers or schools; and Singapore, the self-assessment model for school excellence, with a common assessment at end of primary school to determine secondary school pathways. These demonstrate different models through which systems and schools seek to document, track and improve student learning.

England, which has had shifting forms of assessment at all levels of schooling for student certification and accountability assessments, will implement new statutory assessments for Key stages 1 and 2 over 2018 to 2019. These assessments will inform both summative reporting for students of overall curriculum achievement and accountability monitoring of school quality. Curriculum in England is divided into six stages, not directly related to individual year levels, with Key stage 1 curriculum goals expected to be achieved (and assessed) by the end of Year 2 and Key stage 2 curriculum goals by the end of Year 6. The Key stage 1 accountability end-of-stage summative assessments will be implemented by teachers, against descriptive assessment frameworks for English, Mathematics and Science based on their own assessments and in line with their school's assessment policy (Standards and Testing Agency [STA], 2018a). It is noted that given the changing focus, standards and outcomes are not comparable with those of earlier years.

Guidance indicates teacher judgements are to be based on a "broad range of evidence" from "day-to-day work in the classroom", and "from work in subjects other than the one being assessed, although a pupil's

work in that subject alone may provide sufficient evidence to support the judgement". One piece of student's work may be used for multiple statements of achieved outcomes (STA, 2018b, p. 2). Additional guidance is provided for students with special needs, including options to assess students in "an equivalent way", using teacher "discretion" (STA, 2018b, p. 2).

Key stage 2 accountability measures consist of teacher assessments against similar assessment frameworks with external tests in English grammar, punctuation and spelling papers 1 and 2, English reading, Mathematics papers 1, 2 and 3 (STA Guidance, n.d.) to be completed by all students over four days. Teacher assessment also occurs in science. Of interest, the available Mathematics test practice example is a paper and pencil Arithmetic test requiring students to provide a response to calculations (without calculator) on the test sheet which provides working space.

Statements are of the form "[t]he pupil can …". Moderation of teacher judgement within and across schools is encouraged, not only to attain consistency of judgement but also as a "valuable opportunity for professional development" (STA Guidance, n.d., p. 3). Accountability is met through external validation of a sample of 25 per cent of schools each year to "ensure that [outcomes] are consistent with national standards. It is a collaborative process between schools and local authority moderators" (p. 3). School outcomes are provided electronically to their local education authority, but only by state-funded schools, not private institutions (Department for Education, 2018a, 2018b). While Key stage 1 outcomes do not appear to be reported in a form suitable for use for comparison of schools or construction of league tables, they do form the basis for calculations of student growth to Key stage 2. League tables for Key stage 2 outcomes are published by newspapers (see, e.g., https://www.telegraph.co.uk/education/0/primary-school-league-tables-2017-compare-top-1000-schools). Thus, new directions in accountability assessments in England privilege teacher classroom assessments as well as external tests, providing opportunity for building teachers' professional assessment knowledge, but in the context of considerable comparative publication of performance.

Testing in Japanese schools, administered at the national, metropolitan and prefectural levels of educational governance (38 out of 47 prefectures conduct their own tests), serves as another direction in system-wide assessment (Takayama & Lingard, 2018). National census testing of all students in years 6 and 9 in Japanese and maths was introduced in Japan in 2007 in response to concerns about standards and in an attempt by the national ministry to reassert some central control after a period of decentralisation. What is significant in the Japanese context is the ongoing significance of what are called "instructional advisors" (*shidoshuji*) in all three levels of the education bureaucracy and their roles in relation to the three layers of testing. In many cases, these instructional advisors, experienced and outstanding classroom practitioners, oversee the whole operation of testing, including the design of testing and the construction of test items. No psychometricians and statisticians are involved. Pedagogical relevance appears then to be prioritised over technical validity and reliability in respect of these tests and is a justification for the expensive census nature of testing. Instructional advisors are also involved in compiling result reports to schools and classroom teachers, which detail item-by-item student response patterns and suggest pedagogical interventions to rectify common errors. At all three levels of administration, Japan takes a very cautious approach to the publication of test results to avoid any stigmatisation of underperforming schools. The Ministry of Education only publicises prefectural average test scores, while strongly discouraging

prefectural and municipal boards of education from releasing individual school and school board level data. This is usage of testing geared towards informing and improving teachers' pedagogical practices and a mode of testing informed deeply by teacher professional knowledge and curriculum. Here, teachers still control the field of judgement, albeit mediated by testing. Nonetheless, there is now some tension across the system between support for tests constructed by instructional advisors and for tests created by psychometricians and framed by test theory. There has also been some pressure from Treasury, because of the substantial cost of census testing, for a move to sample testing used simply for accountability purposes, rather than as support for teacher practices (Takayama & Lingard, 2018). However, it seems certain that such a move would be strongly contested by instructional advisors and teachers. It is important to note that the census nature of testing and the heavy involvement of instructional advisors in test item construction in Japan are justified in terms of the resultant support for teachers and the perceived purposes of the tests, namely the improvement of teaching and learning.

Tan (2019) has recently attempted to understand and provide a research-based account of the high-performing education systems in Singapore, Shanghai City and Hong Kong. Here, high-performing is defined in terms of outstanding results on international large-scale assessments (ILSAs), namely the OECD's PISA and the IEA's TIMSS and PIRLS. Tan argues that the success of these systems is a result of a systematic, holistic approach to schooling consisting of a range of complementary policies, what she calls "educational harmonisation". She concedes that "a global testing culture" (p. 67) has framed an exam and test-driven environment in the schooling systems of these three "Confucian heritage cultures". However, and this is very significant, she demonstrates how the three systems in question have also sought to ameliorate this testing culture to a considerable extent and instead give emphasis to a more holistic approach to schooling. She suggests recent reforms in the three systems have sought "to shift from a narrow focus on high-stakes exams to a more inclusive conception of performance" (p. 67). Drawing on Hargreaves and Shirley's (2009, 2012) framework of a Fourth Way of educational reform, Tan shows how there is important alignment across the "pillars of purpose", "principles of professionalism" and "catalysts of coherence". She then depicts schooling in Singapore as "student-centred and values driven"; schooling in Shanghai as focused on quality; and that in Hong Kong as emphasising "learning to learn". Overall, her argument is that these schooling systems, while still utilising testing, are actually moving away from the global trend of standardisation, and test-based modes of accountability. She also stresses the significance of the impact of Confucian values in these systems. She thus argues that direction of policy in the three systems desires much more than "test-taking abilities", but rather "a more comprehensive and learner-centric form of teaching and learning" (p. 93). Tan argues persuasively that the Confucian concept of harmonisation is significant in ensuring coherence and alignment in policy frames in the three system and joins together seeming contradictions such as that between a focus on good test results and critical thinking. While acknowledging that these are geographically-concentrated schooling systems and are framed by different cultures and histories, the significance for the Australian context lies in the importance of policy alignment and in the move away from standardisation with a narrow focus on testing. Testing in these high-performing education systems is now simply one policy element complemented by a range of coherent policies and a stress on holistic, student-centred education. Further, these system-level approaches

provide both opportunity for and emphasis on professional roles for teachers and professional responsibilities in assessment practices.

Further developments: Rich accountabilities

Accountability in many schooling systems most often works through sample rather than census testing. Japan, as discussed, undertakes census national testing, on the justification that the information derived from these tests, created by instructional advisors inside the Ministry, is to assist teachers modify their pedagogical practices to improve student learning on the basis of the evidence derived from the test (Takayama & Lingard, 2018). In Japan, it is acknowledged that if testing were to be used for accountability purposes, it would be of the sample kind; national census testing in Japan is not used for accountability purposes. The influential international large-scale assessments, the OECD's PISA and the IEA's TIMSS and PIRLS, are also of a sample kind and sometimes used for system accountability purposes.

There is an emerging research literature that seeks to rethink what educational accountability might look like as an alternative to the reductive effects of the top-down, test-based mode. This is an argument that says accountability is necessary but needs to be reconceptualised (Lingard, 2009). Sahlberg (2010, p. 53) argues that, "[r]ather than insisting on abolishing school accountability systems, there is a need for new type of accountability policies that balance qualitative with quantitative measures and build on mutual accountability, professional responsibility and trust".

These alternative modes are referred to variously but can be grouped under the category of rich or intelligent modes of accountability. These modes seek a balance between accountability defined as being held to account and giving an account. They also seek to re-instantiate trust in the professional work of schools and teachers. Ranson (2003) argued that until the 1980s the dominant mode of accountability in schooling was a professional one that lacked a complementary public mode, while more recent public modes of accountability deny professional accountability and consequently mistrust schools and teachers. Writing about the successful Finnish system of schooling, Sahlberg (2011) stresses how that system rejects a mode of accountability based on high stakes testing and instead places learning of the broadest kinds at the centre of schools framed by a substantial trust in teacher professionalism. Rich and intelligent modes of accountability in education assert the need for trust in principals and teachers.

Lingard, Baroutsis, and Sellar (2014) argue that rich accountability in schooling needs to be multilateral, multidirectional and mutual. This is a challenge to the unidirectional, top-down character of the test-based mode and its reliance on a single measure of learning. Multidirectionality here refers to accountability of schools to systems, but also importantly of the system to schools, and adds a two-way construction of accountability and responsibility between schools and their communities. This might be seen as a more democratic mode of educational accountability (Biesta, 2004). The model developed by Lingard and colleagues has been derived from an ARC Linkage project involving collaboration between the researchers and the Department of Education in Queensland and is based on close work with schools, principals, teachers, students and community members in a regional part of Queensland (see also Lingard, Sellar, & Lewis, 2017; Lingard, Creagh, & Vass, 2016).

Darling-Hammond, Wilhoit, and Pittenger (2014) offer a model of rich accountability that has these features and which they call "genuine accountability". This rich mode of accountability suggests that consideration has to be given simultaneously in schooling systems to inputs, processes, and outcomes, with different accountability responsibilities situated at classroom, school and system levels, with an emphasis on the relationality between them. At the core of this mode of accountability are meaningful learning, professional capacity building and resource accountability framed by quality and social justice concerns. Darling-Hammond and her colleagues argue that meaningful learning demands a variety of measures of learning that are not restrictive and narrowing of learning and what is learnt and thus place emphasis on multiple and alternative measures of performance (e.g. portfolio assessment). They also argue that the system and schools are responsible for continuing to build the professional capacity of teachers and principals. With the element of resource accountability, they suggest that the system (and policy makers and politicians) need to be held accountable for providing the necessary resources of all kinds to ensure school and teachers can achieve what is expected of them and to overcome any barriers to learning, for example, in schools serving disadvantaged communities. This is the concept "opportunity to learn standards", a vertical bottom-up construction of accountability.

In documenting a model of rich accountability in education, Lingard and colleagues (2014) suggest it needs to be multilateral, involving all stakeholders, and multidirectional, that is, systems to schools and schools to systems and community to schools and schools to communities. Both these multilateral and multidirectional features must work in inclusive, dialogical and reciprocal ways. In terms of the learning aspect of rich accountabilities, Lingard and colleagues stress that the emphasis must be on the learning of all students and learning needs to be defined and measured across multiple domains and in multiple ways; this is a stance echoed in the work of Darling-Hammond et al. (2014). The final elements of rich accountability in education relate to what data are collected and how they are interpreted. There is a question not only what data are collected but also for what purposes; this would include data of multiple kinds, both quantitative and qualitative (e.g. narratives). There is also a question concerning the interpretation of data. This mode of rich accountability sees schools and the systems both held to account, and simultaneously being enabled to provide accounts of their achievements through multiple kinds of data. The pressing question is how to instigate rich educative modes of accountability in schooling systems without intensifying accountability demands on schools and teachers.

Summary

As the Chief Director of Ofsted noted in England in 2017—following school visits and collection of empirical data noting the impact of national assessments in schools (including narrowing the curriculum)—"how easy it is [for school principals] to focus on the performance of the school and lose sight of the pupil" (Spielman, 2017). Spielman identified the situation where what is tested may (inadvertently) become the curriculum.

> It seems unlikely that any school has prioritised testing over the curriculum as a deliberate choice. It is likely that, in some quarters, testing has come inadvertently to mean the curriculum in its entirety. If it is true that curriculum knowledge has weakened across the sector over time, it would explain why there has been a merging of the concepts of testing and the curriculum.

Institute for Learning Sciences and Teacher Education 2018

In general, the impact of accountability testing has been to direct educational policy affecting schools and teacher practices and implementation of curricula, as well as "student learning and experiences of school" (Lingard, Martino, & Rezai-Rashti, 2013).

Overall, research on test-based accountability systems worldwide provides more evidence, while limited in scope and nature, of unintended, generally negative, consequences of test-based accountabilities intended to promote system improvements and limited positive effects on student learning improvement and wellbeing. The research also contrasts use of data, and the nature of data, for the purpose of improving teaching and individual student learning with system monitoring. More recent developments, from England to Asian nations, are placing increased recognition of teachers' professional roles and assessment judgements and the contexts within which they are working.

Review of Australian research literature on educational accountability, improvement and NAPLAN[1]

Introduction

Australian research on the value, use and impact of NAPLAN covers a variety of issues of national relevance, though it is sometimes limited to a single state or small selection of states; where appropriate, attention is drawn to the specific state or states involved. The research has taken place at different times in the decade since NAPLAN was first implemented in 2008, and much of it refers to the early years of NAPLAN implementation. As the response to NAPLAN appears to be evolving, it is important to note not just where the research occurred but when.

Published research on NAPLAN sometimes refers to similar testing and its effects in other countries, often with reference to "high-stakes accountability". However, while it is possible to draw some parallels between testing in Australia and other countries, there are substantial contextual differences that complicate such comparisons. Especially, there is an issue of just how "high-stakes" the accountability is and who is affected by it. The implications of the international research for Queensland and Australia are not self-evident and require separate analysis. One relevant reference which analyses implications for

---

[1] The research covered in this review was identified through database searches for publications referring to NAPLAN, follow-up of further references within those publications, and reference lists generated by the Research Team. Articles or reports offering personal commentaries or viewpoints were not in general included, unless they offered a particularly cogent comment or interpretation on the research evidence. The selected articles and reports needed to be based on data collected by the author(s), or to collate or summarise the data of other reports or offer particular insights relevant to matters identified in those reports, and specific to NAPLAN. The research findings are reported without reference to philosophical or sociological interpretations that may have been offered by researchers to contextualise their findings.

The sections of the review of Australian literature address respectively research findings related to the way NAPLAN is being interpreted and used, research findings on the impacts of NAPLAN; and reports on views about the value and future of NAPLAN. A thematic or topic approach is taken, with the findings of articles and reports disaggregated so that each topic can be viewed across all the relevant articles and reports. Details of research focus, year that data were collected, target groups and methodology are summarised, in general, when an article or report is first mentioned. These details are not usually repeated on subsequent mentions of that article or report.

Australia from U.K. and U.S. research is Lobascher (2011). Here, the focus is on the Australian research, specifically concerning the value, use and impact of NAPLAN.

Published research on NAPLAN is not comprehensive. Each piece of research tends to focus on particular issues of interest to the researchers or as commissioned. There has been no large-scale evaluation of NAPLAN, such as a fully-fledged study of its validity, uses and consequences, as might be expected of any large-scale testing program (Joint Committee, 2014; Thompson, Adie, & Klenowski, 2018). There also are no longitudinal studies that could trace its uses and consequences over time, particularly its long-term effects on students as they progress through school. The research has consisted mainly of: large-scale surveys devoted to particular issues (such as student wellbeing) or particular groups (such as principals, teachers, parents or students); and small-scale case studies of schools, teachers or students (typically as Australian Research Council funded research projects or as higher degree theses). While the corpus of research is therefore somewhat limited in its coverage and style, there is nevertheless a substantial degree of consistency in the findings that lend veracity to the conclusions that can be drawn from it.

General comments on research on NAPLAN

It would seem inevitable that an enterprise as prominent on the national education landscape as NAPLAN should have consequences for Australian schooling, especially through providing measures that allow comparisons between different educational entities and students. There is a clear intention in policies relating to NAPLAN that it should lead to improvement in teaching and learning. It is not surprising then that NAPLAN is shaping changes in perceptions, activities and relationships among the various participants in schooling—administrators, principals, teachers, parents, students—and reconstructing the aims and practice of education (Gorur, 2016; Hardy, 2015a, 2015b). Whether such changes are actually leading to improvement in school practice and student learning has been a subject of much debate (Lingard, Thompson, & Sellar, 2016).

Two different types of consequences of NAPLAN can be identified in the research literature: consequences that are intended and considered to be positive or desirable; and consequences that are unintended and considered to be negative or undesirable—some unintended consequences could be positive or desirable, but these are not common. Consequences can include ways in which NAPLAN is understood, interpreted and used, as well as ways in which NAPLAN impacts the quality and outcomes of schooling, the nature of the school experience for all participants, student learning outcomes, and student health and wellbeing. Unintended (negative or undesirable) consequences have received considerable attention in public (media) discussion and appear more prominently in the research literature than desired (positive or desirable) consequences.

Lingard, Thompson, and Sellar (2016) note a long list of unintended consequences recorded in the research literature, such as performativity emphases, curriculum and pedagogy narrowing and compromise, teacher and student stress and anxiety, inappropriate interpretation and use of test data, and inequitable treatment of students. There are some examples in the research literature of positive consequences, such as collaborative use of data, triangulation of test data with other school-based data, and successful improvement in student learning in broad curriculum aims (Brennan et al., 2016; Hardy, 2014, 2017; Harris et al., 2013; Kerkham & Comber, 2016; Thompson, 2016). Of some importance is that recent Queensland

research has reported substantial unhappiness among teachers (QTU, 2018) and parents (Matters, 2018) concerning the uses and impacts of NAPLAN.

Limits of NAPLAN data

Wu (2016) writes about the technical characteristics of the NAPLAN tests and data, from the perspective of an expert in statistics and psychometrics and offers a cautionary tale. She bases her calculations on information contained in the 2008 technical report to calculate the measurement error associated with the numeracy test (with reported reliability of 0.87) and estimates a confidence interval of 78 points for individual students at a confidence level of 90 per cent.[2,3] Student NAPLAN performance level is reported across 10 consecutive achievement bands from Years 3 to 9, with six bands representing the achievement range at each Year level of testing. A reliability coefficient of 0.87, and confidence interval of 78 points, translates into a possible range across three bands; that is, it is not possible to be certain of student performance ability with (at least) one band either way. This can lead to over-interpretation of the level of precision of the data. Wu (2016) says: "In summary, we would say that a NAPLAN test only provides an indicative level of 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" (p. 23).

Wu (2016) further calculates the potential measurement error for school Year-level cohorts of 50 students or fewer (typical of many schools). Based on the same assumptions as the previous calculations, for a group of 50 students the 90 per cent confidence interval would be 32 NAPLAN points, an average difference of about three numeracy test items, indicating possibly large "natural" fluctuations from year to year. She concludes that for this, and other reasons, the tests "can inform us about performance of large groups of students, but not tell us a great deal about individual students or schools" (p. 28).

There are some implications of these calculations. A recognisable growth in student performance would require a difference in scale scores of at least 156 points (2x78) at the 90 per cent level of confidence; the same goes for distinguishing differences between students in their levels of performance. A recognisable difference between schools of 50 or fewer students in a year level would require a difference of at least 64 points at the 90 per cent level of confidence. These values affect the veracity with which league tables can be interpreted.

How is NAPLAN being interpreted and used? Perceptions of the purpose of NAPLAN

Principal and teacher perceptions of the purpose of NAPLAN

A major large-scale study of principal and teacher views on NAPLAN was conducted by Dulfer, Polesel, and Rice (2012). In May 2012, an electronic survey was sent to all members of the Australian Education Union

---

[2] The more typical 95 per cent confidence interval would be even wider.

[3] In 2017, the NAPLAN Technical Report shows a Coefficient Alpha of 0.87 for Year 3 numeracy, though for other years it is slightly higher at an average of 0.92; for Year 3 Reading, it is 0.84 while the average for other Years in 0.87. The average for all Years in Spelling was 0.91, for Grammar and Punctuation 0.75; and for writing 0.96. These values are sometimes equal to, higher than or lower than the value used by Wu (2016), with 0.87 approximately in the middle of the range. This does not change the general point being made.

Institute for Learning Sciences and Teacher Education 2018

and the Independent Education Unions in each state. Key topics, identified in their previous literature review, were NAPLAN impact on school enrolments, curriculum, teaching, learning, children's health and wellbeing, as well as the purposes of NAPLAN. Responses were obtained from all states and territories, with the greatest response coming from Queensland, whose 3,890 responses accounted for almost half the responses, even though only 20 per cent of targeted personnel were located in Queensland. The state data were weighted to correct for the state differences in response rates, but this made little difference to the data. The validity of the sample was affirmed in terms of its representation of gender, Year levels taught, years of teaching, and experience with NAPLAN.

In terms of perceptions of purpose, the Dulfer et al. (2012) survey found that a large majority of respondents viewed NAPLAN as mainly for the two purposes of *ranking* schools and *policing* schools. Less than half thought it was intended to be useful for parents or to help teachers in teaching students. However, principals had the reverse view, seeing its purpose mainly as a *diagnostic tool* and as a means of *informing parents* of student progress.

Parent and student perceptions of the purpose of NAPLAN

There has been limited research on parent perceptions of the purpose of NAPLAN, and inferences about this need to be drawn from the nature of their comments on other issues. For example, Matters (2018) reports a study of Queensland parent perceptions of NAPLAN, though the terms of reference did not include "purpose". Nevertheless, some data are incidentally relevant. It was found that parents mostly depended on the school for information about NAPLAN (principals, teachers and their children, in descending order), but that they get different messages about the significance of NAPLAN, depending on the school (including that it is nothing special, not a good test, or a waste of time). Parents perceived that the test had become "high-stakes" and that this was not its original purpose. Some, a minority, saw value in NAPLAN for accountability, benchmarking, and benefiting learning, while others thought NAPLAN was inappropriate as a diagnostic test. It is concluded that "the majority of parents surveyed do not fully understand the purpose of NAPLAN and so their ability to fully judge its value and benefits is reduced" (p. 33).

Howell (2016) notes the paucity of evidence on students' own voices about NAPLAN. She conducted a case study of the experiences of 105 students in two Queensland Catholic primary schools. Students did not in general have a clear understanding of the nature and purpose of NAPLAN: "the data suggested that the children experienced the tests within a confusing context of contradictions and dissonances emanating from multiple sources; receiving little, if any, clear and consistent information regarding the purpose and significance of NAPLAN" (p. 564). Some of the student confusion results from the difference between the demands of NAPLAN and the usual helping orientation of their classroom experience, which in turn created a high-stakes interpretation of how the results would be used (including selection into secondary school, and even not being able to get a job if you do badly!). Consequently, Howell suggested there is a "need for unambiguous information about NAPLAN, in language they can understand" (p. 582), together with "authentic opportunities to ask questions about the test and its purposes, with an expectation that their questions will be taken seriously and answered accordingly" (p. 582).

A study by Ng, Wyatt-Smith, and Bartlett (2016) sheds different light on student perceptions, but also reveals limited understanding about the test. Semi-structured interviews were used with 51 Year 5 students in five state schools with low NAPLAN performance and low SES, in urban and rural Queensland. There was a follow-up two years later with a subgroup of 16 students in Year 7 students in the same schools. Some Year 5 students recalled taking a big or long or hard test, but with limited knowledge about the content and purpose of the test. More than a quarter could not remember taking the tests—in general, memories were mainly about test preparation not the testing itself. Knowledge about the test had not improved two years later.

Overall, parents and students have a variety of perceptions of the purpose of NAPLAN, often have limited understanding of it, and would appear to receive many different and confusing messages about it.

Perceptions of the usefulness of NAPLAN

Principal and teacher perceptions of the usefulness of NAPLAN

There is an official expectation that NAPLAN should be useful for schools: "Literacy and numeracy assessments provide rich data about individual student performance and assist teachers to plan learning activities for students. They also enable schools to develop a more objective view about the performance of their students compared to those in other schools and in relation to state-wide standards" (MCEETYA, n.d., p. 1) This statement assumes that NAPLAN can be useful in two ways: to assist teaching by providing "rich data" on student performance; and to illuminate how the school fares in terms of state-wide standards (and presumably therefore as a baseline for improvement). Further arguments in this vein are presented by ACARA (n.d.)[4] and Joseph (2018).

Rogers, Barblett, and Robinson (2018) note the lack of research on the extent to which teachers and parents believe the aims of NAPLAN are being met and whether NAPLAN data is in fact useful in the ways suggested that it should be. Their study was directed at discovering more about teacher and parent perceptions of the usefulness of NAPLAN. They obtained a voluntary sample of 18 Independent schools in Western Australia in 2015, with survey responses from 40 teachers and 345 parents (Years 3 and 5). The teacher survey questions on usefulness showed a somewhat skewed spread of opinions across the range of options, with the average usefulness at about *slightly* (second point, after *not at all*, on the six-point scale). Teachers considered the tests to be highly unrepresentative of the curriculum and to be extremely unfair for some students (in terms of cultural background). This showed a far from enthusiastic view of the usefulness of NAPLAN among teachers.

In the Dulfer et al. (2012) study previously mentioned, about half the teachers thought NAPLAN results were useful or very useful (mostly for identifying surprises and significant areas of weakness, as well as for program and teaching reform). Almost one-third said they did not do more than glance at the data, with the main reasons being that the data were an inadequate representation of student ability or were available too late in the year to be useful. About two-thirds of principals thought the results useful or very

---

[4] http://docs.acara.edu.au/resources/20150424_Reports_supporting_NAPLAN_value.pdf

useful (presumably as a diagnostic tool, since this is what they perceived to be its purpose). Again, the view among teachers (and principals to some extent) was less than enthusiastic.

A subsequent qualitative study, Wyn, Turnbull, and Grimshaw (2014) collected views on the effects of NAPLAN from principals, teachers, parents and students from 16 schools in metropolitan, regional and rural New South Wales and Victoria (29 teachers, 26 parents, and 70 students). The findings were set within other research and literature on the effects of NAPLAN, such as the previous teacher study (Dulfer et al., 2012) with similar conclusions. Principals perceived usefulness to include providing information for individualising learning, identifying students at risk, as a stimulus for up-skilling teachers, and as a benchmark (of progress). On the other hand, they thought that usefulness was compromised by cultural problems with the language of the tests, the multiple-choice format (and guessing), the limited curriculum focus of the tests, the questionable reliability of the tests, and the long delay before results are available. Similar results were found for teachers.

A somewhat different perspective is provided by a small-scale but rich study reported by Pierce and Chick (2011). The study was conducted in the early days of NAPLAN but is nevertheless instructive. The study collected responses from 84 secondary school teachers of English and Mathematics in Victoria and asked about their attitudes to, access to, and use of NAPLAN data. Most teachers (over two-thirds) were positive about the usefulness of NAPLAN data, agreeing that the data are useful for identifying student capabilities and, to a slightly lesser extent, for planning instruction, identifying topics needing attention and identifying student misconceptions. This paints quite a positive picture of teacher expectations that the data will be useful, though the researchers note that the numbers of teachers in the *neutral* category indicate substantial ambivalence, maybe from lack of experience with the data. In fact, teachers who had more direct access to and control over relevant data tended to be more positive about usefulness. There was also a strong view that more use should be made of the data, but that there was a general lack of official encouragement and support for doing so. The Mathematics teachers were more comfortable with using data than the English teachers, and the researchers point to the need for greater attention to training in data literacy (Carey, Grainger, & Christie, 2018; Datnow & Hubbard, 2016), if the perceived usefulness of the data is to be realised in practice. It would seem that data use predisposes teachers to be more positive about NAPLAN but knowing how to use the data is a problem.

The research shows varied views among principals and teachers on the usefulness of NAPLAN, with some revealing greater support than others. One view is that the data could be useful but that teachers lack training in data literacy. On balance, support for the usefulness of NAPLAN is limited or equivocal. Principals see more usefulness than teachers—for identifying student capabilities, topics needing attention, students at risk, progress against benchmarks, and targeted teaching—but both principals and teachers perceive weaknesses in the data that undermine their usefulness, such as: limited representation of ability (lacking richness) and of the curriculum, test format (multiple-choice that allows guessing), inappropriate language and cultural demands for some students, unreliability of the measures, and delay in receiving the results.

Parent and student perceptions of the usefulness of NAPLAN

In 2013, Newspoll conducted a survey of parents on behalf of the Whitlam Institute (Whitlam Institute, 2013), which found that while just over half the parents supported NAPLAN (with one-third against), two-

thirds found NAPLAN information useful (almost all of whom were in favour of NAPLAN). It would seem that, as with teachers, perceived usefulness influences parents' views of NAPLAN.

In the subsequent qualitative study (Wyn et al., 2014), a somewhat different picture emerged. Some 65 per cent of the interviewed parents had reservations or skepticism about the usefulness of NAPLAN but admitted some value in measuring student ability. Their concerns included NAPLAN's unfamiliar seriousness and formality (for students), unfamiliar format and language (for students), and the delay in feedback—somewhat similar concerns to the principals and teachers. The interviewed students generally disliked NAPLAN (especially Year 9 students) and could not understand its purpose.

In the Rogers et al. (2018) survey previously mentioned, parent survey questions on accountability and usefulness showed a fairly even spread of opinions across the range of options, with the average usefulness at about *somewhat* (the middle of the scale). About one-third of parents thought that NAPLAN was a completely invalid and unfair measure of student learning. In open-ended responses, among the most frequent comments from parents was mention of seeing the potential use of NAPLAN results for helping individual students—though severely limited by the delay in obtaining results—and of judging NAPLAN as a poor measure of capability, as a "one-off snap-shot" or when compared with school-based assessments. On the other hand, some parents saw NAPLAN as useful only for comparing schools. There was a strong preference for NAPLAN to be kept low-key. The research did not explore why these views were held.

The Matters (2018) survey of Queensland parents showed substantially negative views on all three dimensions of usefulness: helping teacher work with advanced students; helping teachers work with students who need help; and telling parents where their child needs to improve. Again, we do not know the reasons, though they are presumably related to personal experience.

Uses of NAPLAN data

System uses of NAPLAN data

Lingard and Sellar (2013) studied how the Australian Government in the early years of NAPLAN used NAPLAN data to set performance targets in literacy and numeracy for states in return for monetary rewards. In this sense, the NAPLAN results became "'high-stakes" for states and a catalyst for change. This process of target-setting was part of the *National Partnerships for Literacy and Numeracy* agreement, an initiative in place across the early years of NAPLAN. "The key issues raised by interviewees in relation to this reporting of the National Partnership process included the different levels of 'ambition' reflected in the targets set by each jurisdiction; the diverse nature of targets across jurisdictions; and disparities in the baseline data against which achievement was measured" (p. 642). Targets related to the performance of students against national minimum standards (and mean scores) in reading and numeracy. Targets and methods of data collection differed across the states. Queensland did well by setting modest targets and Victoria did poorly by setting ambitious targets. The selection of target schools also affected the outcome. It was thought that more attention would be given in future to level of performance rather than improvement, and that cynicism engendered by this exercise would lead to increased "gaming" of the system.

The other situation studied by Lingard and Sellar (2013) was that of the Queensland response to poor results on NAPLAN in its first year (2008), in particular the institution of the Masters Review (Masters, 2009a, 2009b). One of the recommendations of Preliminary Advice (Masters, 2009a) from this review was to use NAPLAN assessment materials as a classroom resource—essentially, teaching to the test. The researchers note that the Queensland Premier explained that while "associated explanations [for the poor results] might well be true … there was a political urgency for her to do something, particularly in response to huge and negative media coverage, which was suggesting a problem with Queensland schools … [and] that in this context she had a political problem and had to act, thus commissioning the Masters Review" (p. 647). One consequence was increased monitoring of state schools through the establishment of the Queensland Education Department's *Teaching and Learning Audit* for periodical review of the quality of teaching and learning in each school (noted as a form of goal displacement), together with a blanket target for each school of a three per cent increase in NAPLAN scores each year (both of which, many interviewees in the study thought of dubious value).[5]

Gable and Lingard (2016) note that the Teaching and Learning Audit concluded in 2014 with a change of government. They conclude, however, that these and other public and bureaucratic responses had elevated the importance of NAPLAN to "high stakes" status ("high stakes with real organisational and professional impacts", p. 4), and created new systemic relationships based on managerial test-based accountability rather than professional leadership.

The study that included these conclusions (Gable & Lingard, 2016) focused on interviews in 2013 with principals of two Queensland state primary schools (one socioeconomically advantaged, the other socioeconomically disadvantaged) and their departmental supervisors. These supervisors were newly instituted Assistant Regional Directors—Student Performance (Bloxham, Ehrich, & Iyer, 2015). It was concluded that an institutional rhetoric had been established connecting teaching and learning practices with improvements in learning outcomes (though learning outcomes were now conceived mainly in terms of NAPLAN and My School data). There was an expectation that these data would drive school activity and student learning. Different practices were identified in the two schools: one school, high performing and high SES, employed the recommended process of six-week cycles involving data gathering, data conversations, teacher evaluations, and action planning, supported by strong principal leadership and professional dialogue and development; the other school, low performing and low SES, saw NAPLAN as culturally inappropriate for their students, so that teachers were allowed and encouraged to make their own professional judgments concerning their students' learning needs and progress. In the latter school, problems with literacy and numeracy could only be addressed while dealing with "cultural barriers, poverty and trauma", and performance targets were thought to be meaningless in view of the "shocking complexities" of the students' lives and needs. The first school adapted easily to managerial expectations, whereas the second school experienced a tension between those expectations and the local context. These findings are indicative of some "push back" by schools against demands they consider inappropriate but

---

[5] Dimensions for the audit were: an explicit improvement agenda; analysis and discussion of data; a culture that promotes learning; targeted use of school resources; an expert teaching team; systemic curriculum delivery; differentiated classroom learning; and effective teaching practice.

the extent and consequences of this have not been systematically studied. Neither, too, has the success of the introduction of the institutional structures and controls.

In summary, a fundamental use of NAPLAN at federal and state levels has been the setting of performance targets against benchmarks (national minimum standards and mean score improvement) tied to funding or system-level pressure. This has elevated the importance of NAPLAN to "high-stakes" for schools, principals and teachers, with an expectation that schools can make a difference, learning framed in terms of test performance, and NAPLAN data driving school activity and student learning. While there has been "push back" in some cases against "unreasonable expectations", the consequences have not been systematically studied.

School uses of NAPLAN data

It is surprising how little research has focused on the actual uses of NAPLAN and My School data in schools. It is interesting to know that principals and teachers expect NAPLAN data to be useful (for many, and at least to some extent), but the details about actual use of the data are missing (which data, used in what way, and with what effectiveness). In the studies on usefulness previously reported, it was found that principals and teachers expected NAPLAN data to be used for identifying weak students, identifying student capabilities, planning instruction, identifying topics needing attention, and identifying student misconceptions (Pierce & Chick, 2011; Pierce, Chick, & Gordon, 2013) and for individualising learning, identifying students at risk, up-skilling teachers, and as a benchmark (for measuring success and planning improvement) (Wyn et al., 2014). Matters (2018) found that parents were on-balance neutral on whether they thought schools were using NAPLAN to plan teaching.

There is almost no evidence on whether and how principals and teachers actually do these things. As a result of My School, schools have certainly concentrated on raising their "reputational capital" by focusing on improving subsequent performance on NAPLAN, and general strategies for doing so are known (Hardy, 2015a). But finer considerations of data focus, analysis, interpretation, decision-making and implementation are largely unknown.

The Pierce and Chick (2011) study is one of the few to collect information on teachers' use of NAPLAN data. Survey questions asked who had access to the data, what use was made of the data, whether the teacher received reports based on system data, whether the teacher had access to individual student results, whether the teacher accessed such data, and whether data analysis led to changes in teaching plans. The results are salutary. Only a small proportion of the teachers made use of the data; the main reasons given were lack of access, lack of time, and lack of analytical capability (data literacy).

Limited use of NAPLAN data by teachers was also a finding of Cumming, Wyatt-Smith and Colbert (2016) in their Australian Research Council Discovery Project looking at school and teacher use of NAPLAN data to improve student learning. This finding of limited use was similar to an earlier finding of limited use of the Queensland Aspects of Literacy and Numeracy tests that preceded NAPLAN (Cumming, Wyatt-Smith, Elkins, & Neville, 2006).

Thompson and Mockler (2016) conducted interviews with 13 principals in Western Australia, South Australia and New South Wales in 2013 and 2014, asking among other things about their use of NAPLAN

data. A general finding was that principals liked having the data at their fingertips for detailed analysis in different configurations—allowing identification of program strengths and weaknesses, tracking of students over time, and the focus it encouraged on literacy and numeracy. The study did not examine how they did these things.

One of the striking conclusions that can be drawn from the Thompson and Mockler (2016) study is the extent to which the principals appeared to keep the NAPLAN data to themselves, using it to "drill down" into the data and disaggregate it to unveil the characteristics of the performance of different groups and categories of students, thus keeping an eye on teaching and program success. Hardy (2015a, 2015b), in his case study of a school in northern Queensland, noted a similar tendency, though without much apparent "drilling down". It seems that general practice is to use the NAPLAN data to track performance against the norm for each Year level, with interventions directed to doing better next time.

With the NAPLAN data arriving much later in the year, four months after the testing, their relevance and value are limited (IEUA, 2013; Cormack & Comber, 2013). Much of the value of NAPLAN relies on individual student data, particularly at the item level, but item data are very unreliable. To prevent over-interpretation of the data, a process of triangulation is generally recommended, that is, a process of cross-checking the significance and meaning of the data against other assessments, allowing richer exploration of the data and explanation of anomalies (Renshaw et al., 2013).

Data triangulation, as well as a culture of inquiry, was evidenced in the three schools studied by Hardy (2014a). Information was collected through interviews of principals and teachers in three primary schools in south-east Queensland. Three sets of data were used: NAPLAN data; other external test data (PAT-R and PAT-M); and school assessments in the key learning areas. These data for each student were displayed on a "data wall" (Renshaw et al., 2013) and updated each term. One principal said that this was about embedding data into the mindset of the teachers and asking, "what do you see?". "This engagement cultivated an inquiry-oriented disposition, evident in teachers' recognition of the questioning and discussion occurring within this school about how to improve students' learning … [allowing teachers] … to explore how to improve students' learning in depth and detail" (p. 13). This "educative disposition" was seen as being prompted by the NAPLAN data, while being conscious of limitations of the data, but also how it can highlight deficiencies in particular topics (such as angles) or higher-order thinking (such as reading comprehension). In this case, the principal and teachers were initially responding to external accountability pressures but appropriated that towards a more educative and learning-oriented practice. The resulting "action plans" were not part of this research, but there is mention of ability grouping, building student self-esteem, setting learning targets and personal learning goals. While there is reference to diagnostic uses of data, this would appear to refer to the group rather than the individual student.

Bishop and Bishop (2017) conducted a case study on the use of data walls in a Queensland lower-SES metropolitan secondary school identified at the system level as in urgent need of school renewal. This use of data walls was embedded within implementation of the National School Improvement Tool (ACER, 2012), whose first three components (of nine) are an explicit improvement agenda, analysis and discussion of data, and a culture that promotes learning. Data walls and case management conversations for individual students were an important aspect of building teacher knowledge of students and their learning.

The use of the data walls, based on Sharrat and Fullan (2012), is described in detail with a particular focus on collaborative analysis of the details of the work of selected students (ideally all students) and construction of strategies to support their learning. It is noted that NAPLAN was seen as useful for addressing the needs of some students, but not some groups (such as Indigenous students) and anomalous cases (where school achievement and NAPLAN results differed substantially). As noted earlier, Wyatt-Smith, Adie and Harris (2018) and Wyatt-Smith, Harris and Adie (2018) caution that data walls do not themselves improve learning and that the research evidence suggests they are largely ineffective unless embedded within a broader framework that supports learning (as with Bishop and Bishop, 2017).

An embedded use of data walls is also reported by Singh, Märtsin and Glasswell (2015). They studied the establishment of a collaborative researcher-practitioner knowledge-network (or professional learning community) within a cluster of low-SES Queensland schools. The authors say:

> Our goal in this project was to create a partnership in which professionals with different, but equally valuable, sets of expertise would engage in problem-solving dialogues around student achievement, teacher learning and classroom instruction. These dialogues were designed to be meaningful to all and contextualised to each school site. Student achievement data were gathered using a variety of measures, including diagnostic norm-referenced reading comprehension tests. Teachers met individually with SBRs [school-based researchers acting as coaches and expert resources] to discuss data, examine patterns in student achievement within their classes and consider opportunities for developing innovations that might accelerate student learning. Teachers also met together in Year level teams and as whole school teams, to discuss achievement data for all students in all classes and to reflect on current literacy instruction and share ideas for innovations. The partnership also worked collaboratively on designing, implementing and evaluating the effectiveness of teaching innovations on student learning outcomes. (p. 384)

This connects with international research, which suggests that effective use of data within a school depends on effective leadership, a culture of inquiry, and collaboration (Cumming et al., 2016; Datnow, Park, & Wohlstetter, 2007; Knapp, Copland, & Swinnerton, 2007; Maxwell, in preparation).

Another example of collaborative knowledge networking is provided by Brennan et al. (2016) who studied how a group of low-SES schools in regional Queensland (five high schools and three primary schools) collaborated in response to accountability pressures based on NAPLAN. Principals sought to work together and share knowledge and resources in response to feelings of isolation from central office; their personal relationships were critical for cross-school collaboration. A cultural commitment to student wellbeing and learning was critical. Principals met each other often and shared planning, strategies and policy. Attention was given to teacher collaboration and interaction, especially, sharing of expertise for analysing data, sharing of specialist expertise, sharing in curriculum development, and building professional capacity.

Earlier studies had showed principals largely keeping NAPLAN data to themselves and teachers making little use of the data; more recent research indicates repositioning of NAPLAN data within a broader concept of data to include other indicators of student learning (diagnostic tests and classroom assessments). However, successful use of such data for improving learning is known to require effective

school leadership, a school culture of inquiry, professional collaboration and deliberate building of professional capability.

Student and parent uses of NAPLAN

Parents tend in general to be interested in their child's results on NAPLAN but much less so in their school's results, though that can depend on whether and how the results are conveyed and discussed (Australian Primary Principals Association [APPA], 2013).

The Matters' (2018) study of Queensland parents also reports that parents tend to focus on the individual student reports, but are largely unaware of other reports, such as class and school reports, and to some extent My School. Even so, some find the individual report irrelevant and many do not share it with their child. There was a general belief among parents that they understood the elements of the individual report. However, parents seem to prefer school reports as being more informative and tend to use the NAPLAN report only in reference to level of performance (band level) against the national average, not choosing to make much use of the descriptive information. Some saw the NAPLAN results as confirming their expectations and some saw the results as "inaccurate".

In terms of the political agenda of "school choice" often attached to NAPLAN, Matters (2018) reports that parents strongly disagree that NAPLAN helps them choose a preferred school. In contrast, they strongly agree that schools use results to market themselves.

In the Ng et al. (2016) study, students had largely negative feelings about NAPLAN, but cared about their results if they thought there was a benefit to themselves, and in some cases thought they had learned something from the tests, such as better test skills and content knowledge. However, for both the Year 5 and the Year 7 students, few of them had had discussions with their teachers or their parents about their results, and few of them were aware of the possibilities of using the test results for tailored teaching and self-direction. Students seem in general to be poorly informed about NAPLAN and not drawn into any usage of the results.

Media interpretations and messages about NAPLAN

Mockler (2013) analysed 34 editorials with a focus on My School published in 2010 in Australia's 12 major newspapers (a year after the introduction of My School). She identified three main narratives: of distrust; of choice; and of performance. The *distrust narrative* focused on pillorying those (teachers, unions, governments) who appeared to self-interestedly oppose transparency about school performance (especially through league tables)—school quality needs improvement and teachers are recalcitrant, excluding the public with their obscure educational jargon, and shirking their responsibilities. The *choice narrative* claimed that My School provided important information, otherwise unavailable, for parents to make informed choices based on evidence rather than hearsay, despite recognising that NAPLAN results do not reveal the full picture. The *performance narrative* posed the view that performance—of students, teachers and schools—is best and effectively measured through objective and comparative tests, and that "teaching to the test"—if it is a good and relevant test as NAPLAN must be—is a good thing, and important for the welfare of students and the nation. NAPLAN and My School were represented as important,

therefore, for keeping everyone honest and accountable, as well as providing the foundation for improving the quality of schools and student learning.

Mockler (2016) extended this study by revisiting it in 2013. In this article, she explored three aspects of "'problem framing of NAPLAN in the media" as evidenced in the articles by six key journalists/commentators in 2010 and 2013. These three aspects were: schools or school systems as a problem; teachers and teaching as a problem; and the test as a problem. Journalists differed in their emphases and shifted their positions from 2010 to 2013, tracking changes in the political context and broader debates. It is noted that Queensland's poor results in 2008 were interpreted as a system or government problem, overcome through action by 2013, with attention then shifting to the test as a problem (the development of "gaming the test" through parental withdrawals and encouragement of private tutoring).

As Mockler (2013, 2016) has shown, the media offer particular views about NAPLAN that can change over time and context. Matters (2018) found that 92 per cent of parents surveyed believing that the media give too much attention to NAPLAN results, noting the impact of published league tables.

Impacts of NAPLAN

Quality and outcomes of schooling

NAPLAN and literacy and numeracy outcomes

Evidence of the impact of NAPLAN on student literacy and numeracy across Australia has been demonstrated for the most part in each of the domains of testing. Changes for the period 2008 to 2018 can be examined for Reading, Spelling, Grammar and Punctuation, and Numeracy, and for the period 2011 to 2018 for Writing (later due to the change in Writing focus).

NAPLAN Reading outcomes show statistically significant improvements in Reading for Years 3 and 5 for nearly all states and territories over the period 2008 to 2018, with substantial improvement for Queensland, with the latter influenced by the introduction of the Preparatory Year, an extra year of schooling for Queensland students, in 2008. However, no statistically significant improvements are identified for any state or territory in Year 7 and Year 9 Reading, with the exception of Western Australia in Year 9 Reading.

Queensland's Year 3 Spelling outcomes were also substantially significantly improved over the period 2008 to 2018, with improvement also occurring for Western Australia. In Year 5 Spelling, a number of states, including Queensland, and the Northern Territory showed a statistically significant improvement in Year 5 spelling over this period. Queensland and Western Australia also improved in Year 7 Spelling outcomes, and Western Australia in Year 9.

In Grammar and Punctuation, Queensland recorded a substantial statistically significantly improvement in Year 3, and a statistically significantly improvement in Years 5, 7 and 9. Western Australia also recorded statistically significantly improvements in Year 3, 7 and 9.

Similarly, in Numeracy, Queensland recorded statistically significant improvements in Year 3, 5 and 9, with Western Australia recording improvement in all four years of testing, and a number of other states echoing Queensland's progress in Years 3, 5 and 9.

A topic of interest has been performance in Writing from 2011 to 2018. Across Years 3, 5, 7 and 9 different states and territories have recorded statistically significantly declines in performance. For Queensland, statistically significantly declines occurred for Years 5, 7 and 9, with the decline in Year 7 substantially significant.

Looking at changes occurring in NAPLAN outcomes or performance over consecutive years of testing, the most recent ACARA (preliminary) NAPLAN data (ACARA, 2018f) show no statistically significant improvement in any domain of testing for any state or territory from 2017 to 2018. The only statistically significant change over the two years is a decline in performance for Tasmania and the Australian Capital Territory in Year 5 Writing. Similar stability of outcomes for consecutive years (2015 to 2016, 2016 to 2017) are shown in previous national reports, and for Indigenous students and EALD students. While changes across longer year spans may show some improvements, overall these will be slight and not necessarily statistically significant. It is therefore likely that performance on the NAPLAN domains of testing is plateauing, with the exception of Writing, as predicted by Linn (2000). Thompson (2012, 2013) notes of earlier results, that it seems a lot of effort for limited educative benefit.

The decline in NAPLAN Writing data from 2008 to 2018 in every Australian State and Territory warrants further consideration. ACARA (2017a) shows an increase in the percentage of students who are below national minimum standard in Writing from Year 3 to Year 9. Evidence of "accelerating negative change" holds for each Australian State and Territory (Wyatt-Smith & Jackson, 2016). Understanding the decline in performance and how it might be addressed is important for Australian education

The Australian Writing Survey (AWS) (Wyatt-Smith & Jackson, 2016) was designed to address this significant gap in knowledge to inform policy, research and practice. The primary aim of the survey was to generate information about the practices used by teachers in teaching writing across different curriculum areas and phases of learning. The survey most recently was utilised as part of Queensland Education Horizons 2016 project, *Research partnerships and improvement science: Using data to inform the teaching of writing and assessment* (Wyatt-Smith et al., 2017). The AWS gathered information on teachers' self-reports of how well-prepared they felt to teach writing, based on their initial teacher education, and about the types of professional development that they engaged in with relation to the teaching of writing and classroom practices in assessment. Six hundred Queensland teachers from 55 schools across seven regions responded to the survey. Overall, teachers reported that they felt they had received limited preparation from their ITE training to teach writing and other aspects of literacy. Over 60 per cent of teachers indicated that they were not prepared or minimally prepared to teach reading, writing, handwriting, narrative, persuasive and informative writing, grammar, multimodality and speaking.

The study found greater attention is needed to contextualise the place of writing in all subject areas. One clear finding was a decline in a focus on teaching writing for Years 7 to 10 teachers. Whilst there are policy prioritisations focusing on stages of learning, such as early years and senior schooling, greater emphasis

needs to be placed on the prioritisation of teaching writing in Years 7 to 10. Strategies and resources provided by statutory bodies tend to have a concerted focus from Preparatory to Year 6 in comparison with the middle years of schooling. The study concluded that concentrated focus on the middle years in terms of resourcing and professional development would support teachers' confidence in teaching writing in the classroom.

NAPLAN and literacy and numeracy outcomes for students from specific cohorts

The most recent full NAPLAN National Report on Schooling (ACARA, 2017) notes some slight, though not significant, trends for Indigenous students. However, the average gap between Indigenous and non-Indigenous students remains approximately a two-year lag, that is, for example, Year 5 Indigenous students perform on average at approximately the same level as Year 3 non-Indigenous students. The initial target of halving the gap between Indigenous and non-Indigenous students in literacy and numeracy by 2018 is "not on track" and has so far not been achieved (CADPMC, 2018, p. 58). Results for students who have English as an Additional Language or Dialect (EALD) are more varied. While in some areas initially they have lower scores, they also show improvement in NAPLAN mean scale scores as they progress through Year levels. Details on the achievement of Indigenous students with EALD, compared with other students, are not provided in the national reports on schooling. Similarly, no information is available in the national reports on achievement of students with disability in comparison with other students without disability.

NAPLAN validity

One broad issue is the validity of NAPLAN tests: the question of what the tests actually measure. Surprisingly, annual NAPLAN technical reports do not include an analysis and defence of the validity of NAPLAN as recommended internationally for such tests (Joint Committee, 2014). Validity of the NAPLAN tests is not something that has been much studied. In a strict sense, we do not know what aspects of literacy and numeracy are being assessed and how they relate to broader concepts of literacy and numeracy.

Sometimes this is recognised intuitively. For example, in one study (IEUA, 2013), teachers considered that NAPLAN data gave a "distorted and inaccurate picture" of schools. A more formal point is that NAPLAN is limited in the content and skills it samples and only assesses "fragments" of literacy and numeracy, not the whole of the domains of literacy and numeracy (Harris et al., 2013).

Grasby, Byrne, and Olsen (2015) offer a rare examination of some aspects of test validity in relation to the reading component of NAPLAN. They review some of criticisms of typical high-stakes tests, such as the influence of context and task on reading performance, the influence of task format on numeracy performance, and the complexities and nuances of literacy. They found that NAPLAN reading performance was reasonably well predicted from several alternative and reputable measures of reading, which were poorly related among themselves, indicating that they assessed different aspects of reading. NAPLAN Reading performance was only weakly related to NAPLAN Numeracy performance, since the numeracy test requires reading but also numerical skills. Overall, it is concluded that NAPLAN *Reading* appears to satisfactorily measure a complex of reading skills, that is, has at least partial validity.

Another aspect of validity established in recognised testing standards (Joint Committee, 2014) is related to the use of the test and the consequences of that use. There may be different assessments of validity for different uses. Thompson, Adie, and Klenowski (2018) adopt the argumentative approach to validity of Kane (1992, 2013, 2016), which includes score interpretation and use. They consider the implications of substantially different non-participation rates among the states for drawing inferences on state and school differences and conclude that more needs to be known about the characteristics of non-participation. They note too that, without additional information on "like schools" as presented by My School on the ICSEA measure, it is difficult to make useful interpretations of these data. They suggest that one implication is the need to be more sensitive to the limits of the data, especially what the data mean and to what extent they can be used to make comparisons.

A specific view on the validity of NAPLAN for Indigenous students (and maybe more broadly for EALD students) is provided by Hardy (2013) in reporting a case study of a small rural/remote school in northern Queensland where 85 per cent of the students were Indigenous. Among the school staff, while NAPLAN was rather passively accepted as providing baseline data on student performance in literacy and numeracy, they struggled to recognise its relevance and considered that it was "not testing students' literacy practices in a substantive manner and could be misleading" (p. 73). In particular, they were critical of the comprehension items, seeing the writing task as more valuable than multiple choice, but wanting to use writing tasks more in-situ and diagnostically: seeing how students formulate ideas, revealing their breadth of ideas, seeing how they structure ideas, how they structure text and use spelling in context. Standardised testing was seen as revealing only one facet of student learning, and as not taking into account "the needs of those Indigenous students for whom Standard Australian English was a second (third, or fourth) language" (p. 75). It was suggested that: "[s]uch high-stakes testing practices do not reflect the necessarily situated, engaged, systematic, ongoing, authentic, connected, broad-ranging (individual, small-group and whole-class) literacy teaching practices which characterise more productive/quality literacy practices, particularly for English language/ESL students under challenging material conditions" (p. 76).

Pressures for improved performance (principals, teachers, students)

A persistent commentary from many sources is on the increased pressure that has been placed on principals, teachers and students as a result of NAPLAN. System-level and institutional pressures have been placed on principals for improvement in their school's performance, including targets for improvement, with consequential threats and sanctions (APPA, 2010; Klenowski & Wyatt-Smith, 2012; Lingard, 2010). For example, Lingard and Sellar (2013) describe how this played out through the *Smarter Schools National Partnership for Literacy and Numeracy* (NP). Other initiatives were state, sector or district based.

Pressures for improved performance stem from the public nature of the data together with a focus on comparison. For the states, comparative data are published in the media; for schools, public data are available on My School (and also through league tables constructed by newspapers). These public data affect the reputation of states and schools, who seek to improve their relative position and prevent "reputational damage" (Hardy, 2014b). In this way, NAPLAN has become "high stakes" (Dulfer et al., 2012; Hardy, 2015b; Lingard, 2010; Lobasher, 2011; Polesil, 2014; Wyn et al., 2014) and the main "reputational capital" of states, systems and schools (Lingard & Sellar, 2013). In particular, Dulfer et al. (2012) reported

that about 90 per cent of teachers considered that NAPLAN results could affect the reputation of a school, especially among parents, and consequently affect enrolment of students and teacher morale. Hardy (2015b) also discusses how these pressures were exacerbated by national concerns about raising Australia's performance on international tests such as PISA.

Lingard (2010) discusses how these new pressures on test performance resulted in the shelving of promising Queensland initiatives directed at raising the richness, quality and comparability of teachers' assessments of their students with a focus instead on standardised testing. He describes how this happened through implementation of the *Masters Review* (Masters, 2009a; 2009b), which recommended several actions for teacher professional development, but had a sub-text recommending learning targets (interpreted as performance targets) and the use of NAPLAN materials as a classroom resource. As noted, the Masters Review had been instigated by the Queensland Premier as a response to Queensland's poor performance in 2008 compared to the other states, even though there were defensible reasons for the lower performance (such as younger students). One of the consequences of Queensland's desire to improve its performance relative to the other states was the setting of global targets for improvement on NAPLAN, a form of "using data as central technologies of governance" (p. 650). There was also "increased accountability surveillance through Teaching and Learning Audits and their potential for goal displacement with improved Audit scores becoming the focus of school reform" (p. 651).

Pressures on principals have been transmitted through to teachers. Across all states and across all schools, the pressure to improve results has resulted in general in greater attention to ways of raising performance on NAPLAN and has produced increased workloads for teachers (Dulfer et al., 2012). Hardy (2015a, 2015b) noted that NAPLAN was not only "informing" teachers but "forming" them by encouraging attention to areas for improvement, with a focus on comparisons (with past performance, with other students, with other schools, with state standards). Teachers have sought to become informed about NAPLAN and how best to respond, focusing on how to do better in future. However, this is typically a generalised response focused on improved teaching; Pierce and Chick (2011) discovered very little pressure on teachers to engage with NAPLAN data in detail.

Pressures on schools also come from parents. Thompson (2012) reported that teachers feel pressure from parents for teachers to increase student performance (and that it affects their relationship with parents negatively). However, another study (APPA, 2013) reported that parent pressure on schools and teachers is generally muted (focused on their own child's results), with not many interested in school and teacher performance (as noted above).

Pressures on schools, principals and teachers are also transmitted to students, both explicitly and implicitly (Wyn, 2014). Bousfield and Ragusa (2014), also Ragusa and Bousfield (2017), analysing submissions to the Senate inquiry into the effectiveness of NAPLAN (Senate Standing Committee on Education and Employment [SSCEE], 2014), note reference to "substantial" pressure on students in many submissions: "down the chain of commands (landing with students)" (p. 178). Ward (2012) refers to this as "a results driven domino effect" (p. 112). Matters (2018) also reports that parents perceive pressures are being placed on students to perform well on NAPLAN, with many comments about schools focusing on increasing the percentages of students reaching the top two bands, resulting in giving attention mainly to students in

the middle two bands and more or less ignoring the others. The lack of precision in band placement noted by Wu (2011) is clearly not generally recognised.

An example of this focus on boosting the numbers in the top two bands was Project 600 (or Project U2B, referring to its focus on maximising the number of students in the upper two bands on NAPLAN). This was a technology-enabled on-line individualised program focused on extending students who were already capable and helping them go "from good to great" (Watt, Finger, Smart, & Banjer, 2014).

Pressure on students was explored in the APPA (2013) study. It was found that Year 5 students were impacted more by pressure to perform than Year 3 students were. The reasons for this were given as:

- older students feel pressure from parents and teachers to do better than last time
- some Year 5 students are fearful of the NAPLAN tests based on their previous experience in Year 3
- Year 5 students are more able to understand the importance of the tests and have an awareness of their own ranking and what it means
- Year 5 students have more at stake, such as entry to high school
- Year 5 students are more exposed to school correspondence and media regarding NAPLAN testing. (p. 13)

Teachers also perceive the test itself to be a source of pressure on students. Completion of the test in a 45-50 minute session without a break and without assistance places students under pressure. While this is generally perceived by teachers as problematic, there is some support for a view that this pressure on students is a "good thing" and that it provides students with valuable "life lessons", teaching students about the "real world", though there are also comments about this being an inappropriate use of the test—a form of goal displacement (Thompson, 2013; Wyn, 2014).

Emphasis on testing and improvement

One pervasive effect of the pressures on schools for improved performance on NAPLAN is a tendency towards "performativity" in the culture of schools—an emphasis on performance rather than learning (Lingard, Sellar, & Lewis, 2017). This is described by Hardy (2015a) as "schooling practices characterised by concerns about collecting, analysing and improving numeric data" (p. 1). This involves not just a concentration of effort on improving NAPLAN performance, but a process of adopting a "test driven" approach to teaching and allowing this to trump any concerns that there may be negative consequences from doing so (Hardy, 2016). One of the negative consequences of this emphasis on outcomes rather than learning, based on what we know about incentives for learning, is that it directs student attention to "receiving positive evaluations, or avoiding negative ones" versus making "efforts directed at understanding new ideas and mastering new challenges" (Hatch & Grieshaber, 2002, p. 230) (after Dweck, 1986).

Schools adopt strategies that preference test performance because it is test performance that matters (Gable & Lingard, 2016). A great deal of effort is therefore allocated to practice tests and standardised measures, including the use of commercially available tests (Hardy, 2015b, 2018; Hardy & Lewis, 2017b; Lewis & Hardy, 2017; Klenowski & Wyatt-Smith, 2012; Ragusa & Bousfield, 2017). Comber (2012, p. 127)

Institute for Learning Sciences and Teacher Education 2018

concludes that: "Other criteria for school performance shrink into the background as the NAPLAN data takes centre stage. The dominant texts that come to regulate and reorganise educational practice are now those associated with NAPLAN."

Hardy and Lewis (2017a) note how teachers engage in performativity practices in order to satisfy compliance requirements without believing that the consequences are necessarily desirable. They characterise the conflicts experienced by teachers—"worthless yet important, unnecessary yet indispensable, distracting but beneficial" (p. 682) as Orwellian "doublethink".

> At times, it appeared that the simulacra of the data stories dominated, meaning that the apparent performance of teachers seemed to matter more than their students' actual learning. However, there was also important evidence of challenges to these performative practices, including, vitally, an explicit focus upon whether and how students were learning, and whether the data stories actually contributed to their learning. Such explicit challenges and the focus on evidence of student learning for the sake of student learning, rather than mere representations of student learning, point the way forward to more productive responses to these practices. (p. 683)

NAPLAN preparation

The amount of attention schools give to preparation for NAPLAN is thought to be substantial but has not been systematically researched. One of the few studies to look at this issue is by Dulfer et al. (2012). They note that ACARA itself recommends students should be helped to become familiar with the format and requirements of NAPLAN, but that it is preferable to prepare students through the normal curriculum rather than through excessive preparation.[6] However, they found that teachers believed practice to be important for student comfort, achievement, focus and self-belief. The frequency of testing reported by teachers for the five months before NAPLAN was daily (7%), weekly (39%), monthly (28%), and never (26%), with activity rising as the test approached. Primary teachers reported more practice than secondary teachers. Some teachers commented that too much practice can produce boredom, that students can feel "bullied and harassed", and the result can be low motivation for application during the testing.

Similarly, the APPA (2012) study found wide variation in amount of time spent on NAPLAN preparation reported by principals, who had a typical allocation of 1–3 hours per week in the five weeks prior to NAPLAN testing. Around ten per cent said they begin preparation more than ten weeks prior to the tests.

More recently, Matters (2018) asked parents questions about schools preparing students for NAPLAN and the effects of this preparation on students. The parents perceived that schools spent too much time on preparation for the tests (about two-thirds of parents), were about evenly balanced concerning whether preparation was important, were somewhat uncertain about whether students were well prepared (one-third neutral, though more agreed than disagreed), but mostly agreed that teachers taught to the test (about three-fifths). No association was found between thinking that schools spent too much time on

---

[6] "Excessive test preparation using previous tests is not necessary or useful. NAPLAN tests are not tests students can cram for. Students should continue developing their literacy and numeracy skills through their school curriculum because the tests contain content identical to what is undertaken in regular classroom learning and assessment."
http://www.nap.edu.au/naplan/the-tests

NAPLAN preparation and perceptions of its effects on student motivation. Parents did however think that test preparation sends signals about the importance of the test and that these signals increase student stress and anxiety.

It would seem, as recorded by Dulfer et al. (2012) over six years ago, that there is considerable variation across schools in the amount and kind of attention given to NAPLAN preparation (Matters, 2018). Perceptions of the quality of preparation vary also; Thompson and Mockler (2016) reported principals believing that while other schools employ poor practice, they themselves engaged in "adequate and responsible preparation" (p. 12).

NAPLAN preparation may be especially important, and more sensitive, for Year 3 students. This was an issue considered in the APPA (2012) study but not revisited since. The concern expressed was that Year 3 students lack experience with tests in general, and despite practice may be unsure what to expect, especially since the questions are not typical ones in classroom practice, and these students may lack developmental and emotional maturity to deal with pressures that are inadvertently generated by both their teachers and their parents. A more positive view is offered by Thompson (2013) where 26 per cent of teachers argued that "a positive of NAPLAN was that it had helped students get better at test-taking practices, and the preparation required for the tests modelled desirable attributes such as planning, goal setting and increased engagement" (p. 67). This is another example of goal displacement, as with "life lessons", discussed previously.

Another positive perspective is offered by Anderson (2009) who expresses the view that:

> An opportunity arises [to cater to the diverse needs of students] if we use test items to assist students who have difficulty reading and interpreting mathematical text, to further develop students' thinking skills, and to analyse common errors and misconceptions, frequently presented as alternative solutions in multiple-choice items. One approach to "teaching to the test" is to use NAPLAN items as discussion starters so that students develop number sense, adopt new problem-solving strategies, and build confidence and resilience. (p. 17)

Anderson offers a variety of ways in which this can be done through analysis of errors, thinking strategies, misconceptions, modeling, prototyping, over-generalising, and process-product connection. This is a more nuanced version of teaching to the test, with a focus on important aims in the curriculum. Similar suggestions have been made by Norton (2009), Perso (2009), and Quinnell and Carter (2011.)

Changes in student experience of school

Changes in student experience of assessment

Performativity pressures have impacted on student experience of assessment through the shift towards an emphasis on standardised measures of student achievement, seen both as a basis for preparing students for NAPLAN and as providing more objective information about student capabilities than teacher judgments (Hardy, 2015a, 2015b). Yet the problem with this is that such measures, including NAPLAN itself, are incomplete measures ("fragments") of the learning aims of the school curriculum. If used in an unbalanced way, they give a false reading of what students are learning (Klenowski & Wyatt-Smith, 2012).

Harris et al. (2013) analysed NAPLAN from this perspective and concluded that NAPLAN's limitations included (p. 31):

- NAPLAN's limited coverage of content and skills and the time allocated for sitting the test
- the need for other sources of evidentiary data, including data gathered by teachers in a knowledgeable and principled way, to inform practices that improve learning outcomes
- limitations of NAPLAN as a non-diagnostic assessment procedure for informing improved student outcomes
- validity issues related to attributing students' test scores to school performance and teaching effectiveness
- cultural and linguistic appropriateness and accessibility of NAPLAN's content [for all students].

Similarly, Hardy and Boyle (2011) suggested that NAPLAN could be useful for informing teaching and learning in conjunction with other evidence, but that a focus on the (NAPLAN) measures "erases the complexity of a broader conception of educational practices and ignores the challenges of attending to the diverse needs of real learners, in real time, and in real places" (p. 220). A focus on standardised measures of student learning, by omitting attention to other types of learning outcomes, can render some student capabilities invisible (Hardy & Lewis, 2017b). Renshaw et al. (2013) concluded that students whose needs are not identified by the test measures can "fall through the cracks" and fail to receive adequate attention. The conclusion is that a range of data on student learning, including classroom assessments, is needed to inform teaching and monitor and guide student learning.

Renshaw et al. (2013), from their interview and case studies, found that schools had a restricted view of what counts as data on student learning and contrast this with what the researchers considered effective practice:

> With respect to the data described and exemplars offered, there were none that recorded students' abilities to engage in analysis and evaluation, to apply knowledge and skills to real-life contexts and problem-solving, or to use critical thinking—the so-called 21st century skills. In addition, there were few references to students' skills in communication or related to their affect and social-emotional wellbeing, although one school referred to tracking students with respect to their career aspirations and consequent achievements on their future career paths. Finally, in discussing data, there were few references of the need to take into account or to "read" and interpret data in the contexts of students' access and engagement with learning, their opportunities to learn, and the teaching practices employed. Thus, it is suggested that developing a broader understanding of what counts as data and ensuring that attention is paid to the broader contexts of data gathering so that richer and more nuanced understandings and uses of data can be developed are essential. (p. 14)

> While tests such as NAPLAN will be salient for principals in responding to the external demands of vertical accountability, the effective principal has the capacity to interpret such test results realistically in the context of their school, and to orchestrate whole school assessment and

teaching practices that consider the holistic education of students taking into account the full range of curricular offerings. In short, the effective principal can tell the whole story of learning at the school and identify where improvements need to be made. (p. 15)

Hatch et al. (2002), in the context of earlier state-based testing, remind us that there is a rich and viable history of using child observations in early childhood education, which is being pushed aside by a focus on standardised measures. It is difficult to know whether this is an accurate comment on current practice. Teacher practice in assessment in the context of NAPLAN has not been extensively researched, so details of actual practice are not documented. How current assessment practice impacts on students—their lived experience and how it affects their learning and wellbeing—is also missing.

There is evidence of increased use of tests to monitor and track student learning in literacy and numeracy. In a case study of one school, Hardy (2015a) found constant collection of data through commercially available tests (especially PM Readers and PAT) with constant tracking of student performance against Year level norms. There was constant reference to and discussion of results from such tests "underpinned by the assumption that constant collection would help retain a focus upon broader demands for improved NAPLAN results" (p. 9), but not without contestation by some teachers of the relevance and appropriateness of the data (considering the data as repetitive, and redundant, with a focus on standardised measures). The extent of these kinds of practices has not been systematically studied, but extant evidence suggests that they are widespread.

Changes in student experience of curriculum

There are many reports that increased attention to testing reduces the amount of time available for other aspects of the curriculum (IEUA, 2013; Polesel, Dulfer, & Turnbull, 2012; Swain, Pendergast, & Cumming, 2018; Thompson, 2012, 2013, 2014; Wyn, 2014). This in turn leads to a narrowing of the curriculum as well as reduced diversity of student experiences (Dulfer, 2012; Hardy, 2016; Klenowski, 2016). Such narrowing occurs in general, but especially in relation to what constitutes literacy and numeracy. The curriculum in practice is being shaped by what is tested, which signals what is important (Comber, 2012). The result is an impoverished version of learning outcomes (Gorur, 2016).

It is not known what the long-term consequences of such curriculum narrowing might be. However, there is some evidence that it can lead to decreased student motivation and engagement, and reduced development of creativity, "deep learning" and higher-order thinking (Klenowski & Wyatt-Smith, 2012; Thompson, 2013; Thompson & Harbaugh, 2012, 2013). There is also some evidence of less inclusive and less socially supportive classroom environments (Dulfer et al., 2012; Thompson & Harbaugh, 2012, 2013).

Experiences of students from specific cohorts

Creagh (2016) notes that the classification Language Background other than English (LBOTE) (or students with English as an Additional Language or Dialect [EALD]) covers a broad range of learners with quite diverse learning needs and allows only shallow interpretation of NAPLAN data, which need to be interpreted in the context of language learning, especially second language learning and where students are multilingual but have limited English. Creagh's research is important in its demonstration of how the

statistical category of LBOTE, or EALD, hides substantial performance differences on NAPLAN across the group with potentially negative policy implications.

The testing experiences (before and during testing) of students from specific cohorts (Indigenous students, students with disability, and EALD) have not been much studied apart from the impact of the results on their wellbeing.

An important casualty of the NAPLAN program was the *Embedding Aboriginal and Torres Strait Islander Perspectives in Schooling* (EATSIPS) program. This was a Queensland Government program directed at improving Indigenous students' learning outcomes and at developing greater appreciation of Indigenous cultures among all students. In an evaluation of the program, Vass and Chalmers (2016) found that NAPLAN had undermined the original aims of the program, which had been "appropriated and reconfigured to assist in addressing the literacy and numeracy gap" (p. 139), that is, the difference in NAPLAN performance between Indigenous and no-Indigenous students, thereby disrupting attention to deeper issues. There was little evidence of "meaningfully working towards a deeper pedagogical or curricular engagement with the principles underlying EATSIPS" (p. 146). The researchers considered this to be problematic.

Student self-direction and learning

Very little attention has been given to listening to student voices and studying their experience at first hand. Ng et al. (2016) note from their study of student voices that the possibilities for formative feedback from NAPLAN appear to be mostly unrealised, and that "results have not been used effectively on an individual level to inform learning and provide guidance for improvement for struggling students" (p. 161). Students are largely unaware of any detail concerning their results and NAPLAN appears to have little effect on their learning. The whole agenda of Assessment for Learning (see, e.g., Baird et al., 2004) is missing.

Changes in teachers and teaching

Research evidence identifies that NAPLAN has resulted in changes in pedagogy, largely associated with the performativity orientation. There are differences of opinion about whether these changes are positive or negative, but the majority opinion would be negative (APPA, 2012; Dulfer et al., 2012; Hardy, 2014, 2015a, 2016; Kerkham & Comber, 2016; Klenowski & Wyatt-Smith, 2016). On the positive side is the focus of attention on literacy and numeracy created by NAPLAN; on the negative side are changes towards more teacher-centred, more didactic and less authentic (life-centred) teaching (APPA, 2013; Dulfer et al., 2012; Thompson, 2012). In many cases, these changes lead to less, rather than more, support for those who most need it and to less engagement overall (Thompson, 2012). Many teachers feel frustrated that good pedagogy is compromised (Comber, 2012; Hardy, 2015).

Important insights into the kinds of problems encountered by small rural schools are provided by Cormack and Comber (2013) from their study of a small rural school in South Australia. Such schools have multi-year-level classes and high student mobility between schools, so that NAPLAN data are missing for some students in each class. In a situation where resources are limited, this school resorted to obtaining a US-based reading scheme that offered an easy though one-size-fits-all way of teaching, with consequent lessening of sensitive pedagogical response to students.

Some effects on teachers have been observed. There is evidence of erosion of teacher autonomy and reduced confidence in their own professional judgments (self-efficacy) (Cormack & Comber, 2013; Thompson, 2012, 2013). Teachers also report ethical dilemmas relating to using teaching-to-the-test pedagogy (Comber, 2012), using tactical "cheating" such as giving extra time (Thompson & Cook, 2014), and helping students who become confused and distressed during testing (Comber, 2012; IEUA, 2013). Comber and Cormack (2011) describe how the professional life of school principals has become even more complex and difficult as a result of NAPLAN and its accountability pressures.

There are, however, some examples of "push back" concerning pedagogy, where schools and teachers, with some difficulty, seek to maintain a broadly-based educational program, while attending to the demands of NAPLAN. For example, Kerkham and Comber (2016) report a case study of one high-poverty outer-suburban primary school in South Australia where NAPLAN was seen as "a narrow view of literacy as the practice of content free skills" (p.95) and developed—through leadership, mentoring and collaboration—a student-centred program of literacy learning that respected a broader view of literacy. Another example from a small Catholic school is provided in Kerkham and Nixon (2014).

In the case of the three primary schools in south-east Queensland, studied by Hardy (2014), it was concluded: "That the field of schooling practices is not simply test score oriented *per se* was evident in how teachers elaborated upon the more educational benefits of the testing process, and how the tests could provide useful information to stimulate conversations about how best to effect improved student learning more generally and how they were also recognised as only 'point in time' indicators of student learning" (p. 16).

This was specifically so in the small, mainly Indigenous school in northern Queensland reported in more detail in Hardy (2013). Here there was passive resistance to the NAPLAN data and preference for locally generated data that included commercially available tests (such as PAT, PROBE, Waddington Reading Tests, and South Australian Spelling Test), but also a broad range of in-class data that the teachers generated themselves (such as running records, work samples, semester-long writing and editing tasks, and portfolios), focusing on the language that students used as a basis for further development.

Thompson (2016), from his survey of teachers in WA and SA, noted that, despite the majority view of negative impacts of NAPLAN, there were substantial instances of positive impacts such as closer monitoring of individual students, school-wide coordination of programs, collaboration among teachers and targeted resourcing.

Equity issues

In general, Australian principals and teachers consider that there is more inequity as a consequence of NAPLAN. Inequity is seen to result from:

- reduced attention to lower performing (struggling) students (teacher reallocation of time to test preparation) (Comber, 2012) and giving primary attention to middle-range students ("bubble kids") to raise scores efficiently (Klenowski & Wyatt-Smith, 2012)

- strategic exclusions of lower performing students (Comber, 2012)

- increased use of labelling and grouping in relation to deficits revealed—emphasis on what students can't do (Cormack & Comber, 2013)

- lack of appropriate adaptation for students of difference (minority culture, low SES) who learn differently, and may lack literacy skills and cultural knowledge (Davies, 2012; Dempsey & Conway, 2005)

- inaccessibility of the literacy demands of NAPLAN to students who have not been taught the conventions (its "silent assessors") (Hipwell & Klenowski, 2011)

- limited provision of adaptations for students with disabilities and special needs (Davies, 2012; IEUA, 2013; Mayes & Howell, 2018), especially in comparison with supports provided in teaching, learning and assessment within classrooms (Elliot, Davies, & Cumming, 2016)

- test demands that are inappropriate for students with disabilities given legislative expectations that students with disabilities will participate and be able to participate on the same basis as other students (Cumming, 2012; Cumming & Dickson, 2013)

- expectations of cultural knowledge that Indigenous children cannot be expected to have (Klenowski & Gertz, 2009; Morley, 2011; Wigglesworth, Simpson & Loakes, 2011)

- use of standardised language conventions that mask the language capabilities of the mixed test population of native English speakers, ESL learners and EFL learners in remote Indigenous communities (Harris et al., 2013; Wigglesworth et al., 2011).

Effects on health and wellbeing

Teacher health and wellbeing

The two main factors identified as affecting teacher health and wellbeing have been work intensification (Comber, 2012) and staff morale (Dulfer et al., 2012).

Student health and wellbeing

In their study on the impacts of NAPLAN on students and their families, Wyn, Turnbull, and Grimshaw (2014) offer the following comment:

> The complex interrelationship between student wellbeing and learning is increasingly being acknowledged in educational literature. Over the last 10 years, schools have increasingly focused on creating inclusive and engaging environments, implemented whole school approaches to student (and staff) wellbeing, and acknowledged the role that schools play in addressing anxiety and social exclusion. Across all systems and states in Australia student wellbeing is regarded as an integral aspect of educational policy and practice, because of the strong association between wellbeing and learning. While many students are comfortable with NAPLAN tests, the evidence from this study reveals that NAPLAN tests also contribute significantly to anxiety and to student alienation from learning. (p. 31)

This study was one of the few to concentrate on the effects of NAPLAN on the wellbeing of students—obtaining the perspectives of principals, teachers, parents and students themselves from interviews in sixteen schools across all sectors in New South Wales and Victoria. A large majority of the students disliked NAPLAN, saw it as intrusive and unbeneficial, and thought it should be scrapped. Most felt some stress over the testing, more so if they were struggling with literacy or numeracy, but for most this was a mild and normal reaction. A small number of students experienced sleeplessness and a range of physical reactions, some severe. These or similar findings are found in other research on the impacts of NAPLAN (APPA, 2013; Bousfield & Ragusa, 2014; Dulfer et al., 2012; Howell, 2016; Rogers, Barblett, & Robinson, 2016; SSCEE, 2014).

There is evidence that anxiety about NAPLAN for some students can result in:

- avoidance behaviours (such as truancy, refusal to do the test, reluctance to come to school, and hiding)

- internalising behaviours (such as such as insomnia, dizziness, nausea, sweating, hyperventilation, headaches, stomach aches, crying, and head-banging) (Rice, Dulfer, Polesel, & O'Hanlon, 2016; Rogers et al., 2016; SSCEE, 2014; Wyn et al., 2014).

A majority of parents report that their children are anxious about NAPLAN, which can be interpreted as a fear of doing badly and is considered by parents to be unhelpful (Matters, 2018). Despite this, most parents report that their children have a positive attitude towards NAPLAN and are motivated to do as well as they can (Matters, 2018). Stress and anxiety appear to result mainly from the importance placed on NAPLAN by teachers, parents and the media (Matters, 2018). Some students are anxious about letting their teachers or parents down (Howell, 2016). Others are anxious about what NAPLAN will reveal about them if they do poorly, generating feelings of low self-esteem (Howell, 2016; Rice et al., 2016).

Students can be affected in different ways by NAPLAN. Greater stress and anxiety are experienced by the following kinds of students:

- low performing students (higher performing students report finding the test easy and enjoyable) (Howell, 2016, 2017)

- students from culturally and linguistically diverse communities, those with learning difficulties, and those whose parents have unreasonably high expectations (Rogers et al., 2016)

- students who perceive that the results will be used as a selection device by secondary schools (Howell, 2017)

- students in schools where the high-stakes nature of the tests has been emphasised (and thought they could face "failure", retention or exclusion) (Howell, 2017).

Howell (2017) argues that, contrary to the claim by ACARA that negative experiences on NAPLAN are the fault of teachers conveying stress to their students, "children's experiences of, and responses to, NAPLAN may be a manifestation of their attempts to make sense of NAPLAN within a confusing and at times emotionally charged context" (p. 583). It is also suggested that

the absence of clear and consistent information from adults about NAPLAN's purpose, and the disjuncture between NAPLAN and everyday school life may have led to some children's own constructions of NAPLAN as high-stakes. These constructions bring into question the assumption that because NAPLAN was designed to be low-stakes, children will necessarily experience the test in this way. It is also argued here that what is lacking within current literature is a suitable framework that is cognisant of, and sensitive to, children's own experiences of standardised testing. (p. 583)

Further, the international research on effects of high-stakes testing points out that messages of failure are implicit in the relative comparison process that standardised testing is based on—some always fail (Hursh, 2005; Kohn, 2001; Linn, 2000)—and that poor performance can follow students throughout their schooling. Reinforcement of their "failure" at each subsequent testing can have detrimental effects on their wellbeing (Cumming, Wyatt-Smith, & Colbert, 2016).

Student participation in NAPLAN

All students in Years 3, 5, 7, and 9 are expected to take the NAPLAN tests. ACARA's guidelines indicate that the following students can be formally exempted:

- recently arrived in Australia and with a language non-English speaking background

- having significant intellectual disability and/or significant coexisting conditions

and that students with a disability may apply for appropriate adjustments. (ACARA, 2018e).

In the Dulfer et al. (2012) study, it was noted that the overwhelming majority of students sat the test, though withdrawals varied across schools. Teachers gave the following reasons for their recommending removal of a student: eligibility for exemption (almost 90% of teachers); possible negative effect on student confidence (50%); nothing new would be learned about the student (40%); or the student would not be able to concentrate for that long (30%). Withdrawal was more common in primary schools than secondary schools (by 3 to 1).

When it came to parents' reasons for withdrawal, however, principals and teachers said they thought the most common reasons were (in descending order of frequency): possible negative effect on student confidence; opposition to NAPLAN; nothing new would be learned about the student; absence of family at time of testing; student would not be able to concentrate for that long; the student is too young for formal testing; and it would distract from normal learning. Teachers' reasons for recommending withdrawal and teachers' perceptions of parents' reasons for withdrawal were very similar, with concern for student confidence foremost (after official exemption) in both cases. This seems to relate to students who would find NAPLAN challenging, maybe not for the first time, and the possible distress at experiencing "failure once again". Some children were reported as simply truanting or feigning illness.

Views on the value and future of NAPLAN

Matters (2018) asked parents about the value of NAPLAN for various stakeholders. This was reported in two ways: the percentages of parents who ascribed a zero value (*no value at all*) to each stakeholder group;

and the mean values, on a 10-point scale, ascribed to each stakeholder group by those parents who thought there was some value.

Close to one-half of parents (45%) thought NAPLAN had no value for the people of Queensland; more than one-third thought it had no value for students and parents; more than a quarter thought it had no value for both teachers and the federal government; and more than a fifth thought it had no value for both their school and the Queensland government.

The ratings of value given by the "non-zero" parents were highest for the Queensland government (scale point of 6), declining through value for their school, the federal government, teachers, parents, students, and to the people of Queensland (scale point 2). These figures indicate that parents have, at best, a modest view of the value of NAPLAN, and then essentially for bureaucratic and accountability ends. Matters (2018) reports that parents listed "worst" features more easily and more often than "best" features, and that worst features were typically concrete and local (related to student experiences), while best features were more abstract and systemic. The report concludes that, on balance, the parental view is that NAPLAN lacks worth.

Matters (2018) reports that some parents, apparently a minority, took a pragmatic view that the aims of NAPLAN were reasonable, but the execution was poor. They sought changes that would better satisfy the aims. Thompson, Sellar, and Lingard (2016) offered another perspective. They asked, again pragmatically, whether—given that some schools seem to manage to avoid the worst excesses of performativity pressures—whether it would be useful to explore what schools reporting positive impacts are doing, and whether it would be useful to reappraise what constitutes valid use of test data at classroom, school and system levels.

CHAPTER 3: METHODOLOGY

The aim of Phase 2 of the 2018 NAPLAN Review was to explore the responses of multiple participants in the Queensland education system. The project's focus was primarily to engage with participants who had differing roles, responsibilities and experiences in relation to NAPLAN, and provide all educators with an opportunity for a voice regarding their experiences of NAPLAN. To enact this aim, multiple research tools were created to tailor and cater for specific audiences collecting both quantitative and qualitative data to ensure that all participants were validly represented.

Central to the design of all research tools were the Terms of Reference (ToR), and specifically the issues raised in ToR4 that the Review Phase 2 was to address. As noted, the issues to be examined included, but were not restricted to:

- the value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level

- how Queensland NAPLAN data is utilised, communicated and reported within schools, the broader education system and the community

- expectations, understanding and use of NAPLAN results by students, their families, school leaders and systems, and its importance in accountability and monitoring of student outcomes

- factors affecting NAPLAN participation

- evidence of the impact of NAPLAN on student and staff wellbeing

- the effect of NAPLAN on the ability of teachers to teach the full curriculum, school leaders to progress curriculum and program priorities, and schools to deliver on broader educational objectives

- how NAPLAN affects specific student cohorts, including Aboriginal and/or Torres Strait Islander students

- the differentiated experience of schools and students that participated in NAPLAN Online in 2018

- the impact of NAPLAN on school and system resourcing and

- any undesirable consequences for students, teachers, school leaders, schools and the education system. (DoE, 2018, p. 3).

Surveys items, interview and focus group discussion questions were aligned against these issues and conclusions drawn from the data and analyses in response to ToR4.

The reviewers employed a mixed-methods approach to the study. Online surveys, using the platform SurveyGizmo, were used to ensure the maximum number of participants was able to contribute their views on NAPLAN, while interviews and focus groups with key participants were organised across seven regions to ensure all levels of systems and sectors of education had an opportunity to participate. Both quantitative and qualitative analyses have been undertaken. The timeframe for the staged approach to data collection (Table 3.1) is inclusive of surveys, interviews and focus groups. Table 3.1 includes the number of active participants in the project.

Table 3.1 Timeline of data collection

Data	Participants	Number of participants	Start date	End date
Literature Review	National and international research		31/8/18	17/10/18
School Survey	Registered teachers in Queensland schools	5,814	31/8/18	17/9/18 (18 Days)
Student Survey	Students in Years 3, 4, 5, 6, 7, 8, 9 and 10	2,896	3/9/18	17/9/18 (15 Days)
Organisation Survey	Targeted organisations/associations	4	4/9/18	17/9/18 (14 Days)
Interviews with key participants	21 nominated key participants from relevant associations	21	3/9/18	7/9/2018
Focus Groups				
Regional focus groups (site visits)	Across 7 Regions: Principals, Regional Directors, Assistant Regional Directors, Sector Regional Representatives and key departmental participants	86	10/9/18	20/9/18
Specific focus groups	Queensland Aboriginal and Torres Strait Islander Education and Advisory Committee (QATSIETAC)	8	10/10/2018	10/10/2018
	Deans and Heads of Education in Higher Education	5	09/10/2018	09/10/2018
Total number of Participants for all Focus Groups: 99				

A literature review of both national and international research used to inform the design of the surveys provides a further framework for findings and discussion.

Participants

The School, Student and Organisation Surveys (Cumming, Maxwell, Colbert, & Jackson, 2018) were targeted to three specific audiences. The **School Survey** was designed to engage principals, middle management, and teachers (both primary and secondary). The **Student Survey** engaged specifically with students in Years 3 to 10 and the **Organisation Surve**y was designed to target relevant associations across Queensland. All surveys were cross-sectoral with participants in all education sectors invited to provide their views on the role of NAPLAN in school and system improvement.

Ten interviews were conducted with twenty-one interviewees involved with NAPLAN from a system perspective. Interviewees included sector authority CEOs and senior management involved in policy, performance monitoring and improvement; QCAA personnel and Teachers Union representatives.

The regional focus groups took place in the seven Department of Education-identified regions in Queensland (Figure 3.1): South East, Metropolitan, North Coast, Darling Downs, Central Queensland, North Queensland and Far North Queensland. Regional focus groups included principals, regional directors, assistant regional directors, sector regional representatives and key departmental participants. Additionally, specific focus groups were also conducted, with the Queensland Aboriginal and Torres Strait Islander Education and Training Advisory Committee (QATSIETAC) and Deans and Heads of Education in higher education.

Figure 3.1: Seven State Primary and Secondary School Regions of Queensland

Design of surveys, interviews and focus groups: Terms of Reference

The School Survey was designed to ensure that multiple roles were captured with two pathways designed dependent on the role of the participant. Leaders in the School Survey such as principals, deputy principals and curriculum leaders answered questions relating to their experiences with NAPLAN commensurate with their roles in their school. Teachers answered questions aligned with the leaders' questions contextualised with their role in the school.

Items in the School Survey reflected an extension of a survey developed for a previous project, *School and Teacher Use of NAPLAN Data for Student Learning Improvement* (externally-funded Australian Research Council Discovery Grant DP110104319) and informed by research literature at that time, subsequent literature and ToR 4 issues.

Student Survey questions were also aligned with ToR 4 issues and designed to engage multiple age groups. Participants were given multiple question pathways dependent on experience with NAPLAN, NAPLAN Online in 2018, and Year level.

The Organisation Survey items were aligned with the School Survey. Questions relating to demographic information were excluded, as were key questions relating to specific school experiences (see Figure 3.2 below).

Figure 3.2: Outline of the School Student and Organisation Survey
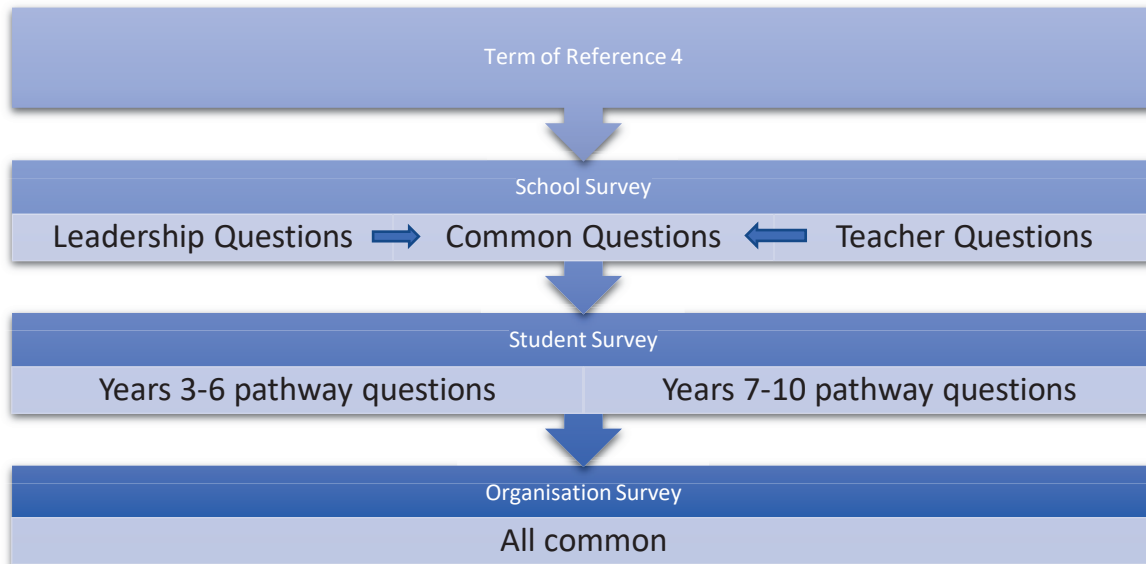
Collaboration with Department of Education key personnel occurred through formal meetings, by email and by phone for the purposes of ensuring item specificity, clarity and scope and to confirm that the survey items and constructs addressed appropriate issues.

Survey distribution and collection

The School Survey was distributed to all 70,233 registered Queensland teachers via emails sent by the Queensland College of Teachers (QCT). Participants were asked to identify their school name, enabling merging of participant responses with demographic school data (level of schooling, region, geolocation, sector, ICSEA) from data provided by the QCT. School Survey participants were therefore not required to provide such information about their school. All surveys were anonymous for individuals and school names have been removed from the final data file.

To ensure adherence to ethical guidelines, a link to the Student Survey was sent to schools via the relevant sector representatives with an accompanying letter to explain the context of the Survey. Sector representatives outlined that the Student Survey link, with accompanying information, was to be embedded as part of the school newsletter or email to raise parent/carer awareness. The Student Survey required both parental/carer permission and student permission. Student Surveys were anonymous, no record of student name, school or sector was collected. Students were asked to provide their current Year level, their gender, and whether they identified as an Indigenous person.

The Organisation Survey was emailed to key stakeholders nominated by the Queensland Department of Education.

As highlighted in Table 3.1 earlier, the School Survey was open for 18 days, the Student Survey for 15 days and the Organisation Survey for 14 days.

Online Delivery: SurveyGizmo

The online platform used to deliver the surveys was Survey Gizmo, an online survey software tool. This software was selected due to the following features:

- customisation options provided for survey design, including question type, question logic as well as overall visual design
- viewing options for multiple platforms including desktops, laptops, tablets and phones
- options for email link delivery, and the option to save and continue links that allow respondents to save survey progress for later completion
- ability to generate ongoing reports on responses and completion rates and
- security measures that ensure data are secure.

Communication

To maintain maximum engagement, a variety of communication strategies was employed to promote completion of the surveys; education authorities were instrumental in supporting the delivery of this communication through a variety of platforms outlined below:

- Queensland Education Minister's media release
- ILSTE Twitter posts
- QCT emailing of teachers
- Queensland Education Minister email to teachers
- Department of Education Facebook and Twitter posts
- Promotion through sector parent group websites.

The communication strategy contributed to surges in survey completions and was integral to the success in participation rates.

School, Student and Organisation Survey

School Survey

The 5,814 responses received for the School Sector represented all sectors, regions and school types with an overall response rate of 8.3 per cent. This is a substantial number of responses on which to base conclusions.

In order to check the representativeness of the responses, several analyses were conducted. Participation was analysed across sectors, school types, regions, geolocation, roles, gender, teaching experience, and teaching level. These analyses demonstrate a well-balanced response across all these categories.

Apparent discrepancies in participation were balanced by other considerations. For example, the representation of primary schools was lower (74% of schools) than for secondary and combined schools (average 95% of schools); however, the response rates from primary teachers and secondary teachers

(7.3% and 5.6% respectively) showed a slight trend towards primary teachers but were not very different from the overall response rate (8.3%). The percentages of respondents indicating that they teach in a primary school or a secondary school also match closely the percentages in the population for these two categories as identified by the QCT.

Overall, response rates show excellent representation of sectors, school types, regions, geolocation, role, gender, experience and teaching level, with no evident bias to affect data and interpretation. It is, of course, not possible to know whether non-respondents would offer different judgements from those who participated in the School Survey. However, to the extent that the sample of respondents covers all descriptive categories of the state population of registered teachers and their schools, and does so in a substantially representative way, some confidence can be held in the generality of the data and the findings.

School Survey: Completion statistics

The use of SurveyGizmo software supported ongoing review of reports on response types and completion rates for the online survey. This indicated that, of the potential participants who clicked on the survey link, the 5,814 responses had an average completion rate of 388 per day. The number of completions was communicated regularly to the sector representatives to ensure up to date information was provided and opportunities for promotion were utilised.

The response activity on the survey webpage was analysed to identify factors that will assist future planning strategies for survey completion. This analysis provides information about the platform and operating system used to complete the survey. Over 62.7 per cent of the responses were completed using a Windows desktop computer or laptop, 8.8 per cent used a Mac desktop computer or laptop while the figure for mobile phones and tablets was 28.5 per cent . While completion on a desktop would be expected, mobile phones appear to be an emerging platform for people to complete the survey. This information highlights the importance of survey design being suitable for use on mobile devices to contribute to a higher completion rate.

Student Survey

The Student Survey was targeted specifically for students in Years 3 to Year 10, with multiple pathways designed to ensure age-appropriate items commensurate with experience with completing NAPLAN and school phase. The Student Survey generated 2,896 responses, evenly distributed across all Year levels, with 733 in Years 3 and 4, 762 in Years 5 and 6, 762 in Years 7 and 8, and 639 in Years 9 and 10. One hundred and twenty-five students responding to the Student Survey had not completed NAPLAN so did not complete all survey items. These students were asked if they knew the reasons for nonparticipation in NAPLAN, and then thanked for their involvement before being exited from the survey.

Organisation Survey

Organisations were provided with opportunity to comment on NAPLAN through a specific Organisation Survey with items similar to those provided on the School Survey. Representatives for four organisations completed the survey, three from professional teacher associations, and one from a special education association.

Institute for Learning Sciences and Teacher Education 2018

Interview sessions

Following the guidance provided by the Department of Education, representatives of stakeholder organisations from all sectors were asked to participate in interviews. Such organisations include sector authority CEOs and senior management involved in policy, performance monitoring and improvement; QCAA personnel and the Queensland Teachers Union and Independent Education Union. Ten interview sessions were conducted with 21 persons in total. Interview questions, as discussed earlier, align with issues identified in ToR. 4. Interviews were recorded and de-identified, and written transcripts produced.

Focus groups

Participants were nominated by the Department of Education and Sector heads to contribute to the focus group sessions. Ten focus groups were conducted with 99 persons. Focus groups were organised to ensure engagement with key participants. The focus groups were in four distinct groups:

- Regions—inclusive of principals, curriculum leaders, sector regional heads or acting regional heads; assistant sector heads

- Organisations—inclusive of senior departmental managers involved in policy, performance monitoring and improvement; literacy and numeracy consultants; representatives of the Queensland Principals Associations (inclusive of Primary, Secondary and sector)

- Queensland Aboriginal and Torres Strait Islander Education and Training Advisory Committee (QATSIETAC)

- Deans/Heads of Education in higher education institutions training Queensland teacher education students.

Regional focus group participants were from the Government sector (60.5%), the Catholic sector (32.5%), and Independent schools (7%). They comprised school principals (57%), regional directors (16%), assistant or deputy directors (21%), deputy principals and heads (2.3%), and individuals in other administrative roles (3.5%). Researchers travelled to all seven Queensland regions to conduct the focus groups. Teleconference options were arranged in each region and accessed by several participants.

Focus group discussions followed a semi-structured interview schedule reflecting issues identified in ToR 4 with a brief PowerPoint presentation provided at the commencement of the focus groups to show the overall focus of data collection and as a prompt for discussion. Participants discussed the key issues from the perspective of their role in education and commented on other areas in the system as appropriate to their context. All focus groups sessions ran for approximately one hour.

All participants were voluntary. All focus groups were recorded, de-identified and written transcripts were produced.

Data analysis

Quantitative data

Analyses of the responses to the School Survey used simple descriptive statistics (frequencies and cross-tabulations). Data were analysed for all participants by region, type of school (primary/secondary/combined), role in the school (principal, middle management [deputy principal, head or dean, head of curriculum] and teachers [secondary, primary, learning support and other specialist roles]). Responses to the Organisation and Student Survey were analysed using simple descriptive statistics.

Qualitative data

Interview and focus group transcripts were analysed using NVivo 11 qualitative data analysis software (QSR International, 2015) to examine core themes and issues raised in response to the semi-structured interview and focus group questions.

Qualitative method(s) of data analysis: Interviews and focus groups

Using NVivo 11 (QSR International, 2015), three phases of data analysis were conducted. In the first phase, a new project database (.nvp) was created and the eighteen (n=18) interview and focus group transcripts were imported into NVivo 11 first as "internal sources". Following this importation, word frequency queries and text search queries were conducted using key terms derived from interviewers' field notes and preliminary scanning of imported sources. The initial list of these key terms was collectively formed as part of a group discussion between the interview facilitator and assistive facilitators and included terms such as "accountability", "transparency", "diagnostic", "triangulation", and "media". These initial key terms were then used to create nodes, or coding groups, and a corresponding node hierarchy (e.g. parent node, first-level child node, second-level child node, etc.). This node hierarchy formed the initial coding framework for the remaining phases of data analysis . The imported sources were then auto-coded in accordance with these nodes, via text search queries, and aggregated to reveal the total number of key term references for the sources.

In the second phase of analysis, the auto-coded sources were manually searched, assessed and noted for content-relevance (i.e. context). NVivo 11 software tools including highlighting, coding stripes and memos were used throughout this process. Additionally, relevant key terms derived from this assessment process, such as the text wrapped around key terms (e.g. "accountability agenda"), were then coded in vivo (QSR International, 2015). Following the creation of these new nodes and corresponding node hierarchy, additional text search queries were conducted. These additional coding references were then manually assessed, as described in the first portion of this phase, for content-relevance. All contextually inappropriate or irrelevant codes, from the initial and additional assessment processes within this phase, were manually deleted.

In the third phase of analysis, the extended version of the node hierarchy was examined and the coding categories with the highest coding density—that is, the terms with the highest number of coding references by interview and focus group participants—and/or the most contextually appropriate coding references

were explored. These categories were clustered together to create appropriate themes to guide thematic interpretations of the qualitative data.

The final two focus groups conducted with QATSIETAC and Deans of Education were also guided by these NVivo analyses and interpreted thematically. Discussions of the QATSIETAC focus group were incorporated with discussions from the regional focus groups.

Qualitative method(s) of data analysis: Qualitative School Survey responses

Three thousand four hundred and forty-three (3,443) respondents also provided qualitative commentary on the School Survey. Guided by the method(s) of analysis applied to interview and focus groups data, these comments were also examined using NVivo 11 qualitative data analysis software (QSR International, 2015) and comprised three phases of analysis. In the first phase, qualitative survey responses were preliminarily scanned in excel and separated according to respondent's role(s); namely, principals, deputy principals and heads of curriculum/departments, teachers (secondary, primary, other), and learning support teachers. A new project database (.nvp) was then created and the 3443 teacher responses were imported into NVivo 11 as "internal sources".

Word frequency queries and text search queries were conducted using key terms derived from the preceding preliminary scan of the data. These initial key terms were used to create individual coding categories and a corresponding node hierarchy, which formed the initial coding framework for the remaining phases of data analysis. The imported sources were then auto-coded in accordance with these nodes, via text search queries, and aggregated to reveal the total number of key term references for the sources. In the second phase of analysis, the auto-coded sources were manually searched and assessed for content-relevance (i.e. context). In the third phase of analysis, the final version of the node hierarchy was examined, and the contextually appropriate coded data was extracted, synthesised and interpreted thematically.

Qualitative data analysis: Qualitative Student Survey responses

Three hundred and sixteen qualitative comments were provided on the Student Survey. These were analysed thematically for main topics raised, and coded to create simple tabulations of frequency of occurrence of topics.

Qualitative data analysis: Qualitative Organisation Survey responses

Only two qualitative comments were provided for the Organisation Survey. These are discussed in Chapter 4 as relevant.

Ethics

Ethics for the project was approved internally through Research Services, Department of Education Queensland on the 22nd August 2018. This project was also approved by ACU ethics, registration number 2018–193HE. All survey data were de-identified in line with ethical procedures for this project. Consent for all participants in interview and focus groups was obtained.

CHAPTER 4: FINDINGS AND DISCUSSION

Introduction

The overall focus for the 2018 Queensland NAPLAN Review Phase 2 (Sector and Schools) was to examine use of NAPLAN in the Queensland context, the contribution NAPLAN makes to "enabling students to reach their full potential" and the role NAPLAN "plays in school and system improvement" (DoE, 2018). This focus positions the Review therefore to examine NAPLAN in the context of multiple potential goals: assisting all students to gain excellence in educational outcomes; assisting schools to improve teaching and learning programs to achieve these outcomes; and informing systems to identify and facilitate such performance. The Review's overall focus aligns with ACARA's description (2018c) of multiple purposes for NAPLAN—data for system and school monitoring, to support teaching, learning and school improvement, for mapping student progress and identifying students needing support, and provision of information to parents and the community on their child and child's school—albeit here in reversed order, from monitoring at system and school level to school, student and community purposes. A monitoring system becomes an accountability system when it becomes high-stakes for those involved and is used as a "stick" to "drive" educational improvement, rather than as a valuable source of information.

Historically, as noted in Chapter 2, Australia's literacy and numeracy testing evolved to address the two purposes of individual student achievement and system–school monitoring. Initially, the focus on literacy and numeracy testing at state and territory level was to identify individual students "at risk" and provide diagnostic information for those needing additional support. The goal was to have all, or nearly all, students achieving minimum expectations for literacy and numeracy, as they became defined over time. Implicitly, literacy and numeracy testing was positioned as a means by which such information on achievement could be reliably obtained on a widespread scale.

The national declarations, commencing with the Hobart Declaration (MCEETYA, 1989), reinforced policy focus on individual students, equity, and expectations for all students, but also introduced the policy of national reporting as a measure of overall educational performance. The introduction of the common national test, NAPLAN, in 2008, strengthened the role of literacy and numeracy testing for educational accountability and the "transparency" of schooling outcomes.

Internationally, test-based accountability systems also address two fundamental goals of system monitoring—the "quality" of schooling as well as the realisation of educational excellence for all students, including students disadvantaged due to socio-economic factors and students with disability. For example, national testing in the UK is intended to promote transparency regarding school quality, noting that

national testing and data publication should lead to learning improvement (DfE[UK], 2010). The US *No Child Left Behind* (NCLB) legislation (NCLB, 2002) was initially aimed at promoting learning expectations for all students through increasing state and school accountability for learning of all students. However, system-level accountability became the dominant purpose of NCLB in practice.

Purposes for test-based accountability or monitoring systems are identified as having broader potential educational purposes at the school level including focusing instruction on important content, defining expectations for students, and providing information to schools and teachers about student achievement, including identification of students at risk (Madaus et al., 2009; Russell, Madaus, & Higgins, 2009). These parallel the multiple purposes noted for NAPLAN (ACARA, 2018c). With respect to NAPLAN, caution has been noted regarding the extent to which NAPLAN can provide individual student's diagnostic evidence or enable individual student tracking over time, given issues of test reliability (Wu, 2016). Small schools may similarly not benefit from NAPLAN data as much as schools with larger student bodies, for the same technical reason (Wu, 2016).

Notwithstanding ACARA's ascription (2018c) of multiple purposes to NAPLAN, then, the question becomes the extent to which a single national test-based system is perceived as meeting multiple purposes—addressing individual learning needs and providing useful evidence to inform teaching and learning programs, while also focusing on school and system monitoring as the driver to benefit and improve future teaching and learning. The sections in this Chapter address this question through consideration of the themes underpinning issues identified in Term of Reference 4 to be addressed: Purpose of NAPLAN, Value of NAPLAN, Uses of NAPLAN, NAPLAN and Students from Specific Cohorts, the 2018 NAPLAN Online experience, Impact of NAPLAN and Improvements in NAPLAN.

Purpose of NAPLAN

Review participants in interviews and focus groups identified a range of purposes for NAPLAN from their perspectives, both as a "tool" and as data. They frequently qualified their views of "what NAPLAN is now" with their perceptions of the original purposes of NAPLAN and desirable purposes for NAPLAN. In discussing purposes, participants tended also to discuss *Values* and *Uses of NAPLAN*, which are addressed more fully later.

Participants' viewpoints echoed the range of purposes identified by ACARA (2018c) included benchmarking student performance, whether against standards or against others, "a systematic check", and jurisdictional accountability purposes at several levels including federal and state and territory levels, regional and school levels. While many participants noted both student-focused and accountability purposes, they weighted the two differently; some emphasised the system accountability purposes, and others emphasised student-focused or teaching-focused purposes.

At the system level, references were made to a "health check" on the system. NAPLAN was described as the only "large-scale systematic" and "big" data providing "a snapshot" of literacy and numeracy learning at key junctures, to see "if what we're doing in terms of implementation and support for schools is actually [improving] outcomes for students". This was deemed to be "vital" and "positive". Advantages were seen in the availability of information that was standardised, providing a point of reference for accountability

with capacity to analyse trends in schools. NAPLAN was identified as a "really useful mechanism" to track growth at different levels and for comparison of student cohort progression. Within regional focus groups, discussion mentioned one purpose as enabling comparisons and monitoring of regional school performance within systems.

At school level, the purpose of NAPLAN was seen as provision of information to schools and teachers about curriculum and areas in teaching and learning programs that were strengths or needing to be addressed, and monitoring over time. Further, "identifying where children are at in their learning in order that teachers can help them progress" was seen as a "very, very specific [intended] purpose" at the "school site and at the classroom level", the student-centred focus.

> … if [NAPLAN] was used for its intended purpose it would enable teachers in the classroom to have a judgement, of their students', individual students', work against national standards … fundamentally I thought [NAPLAN] was designed [for diagnostic purposes and] that we would do better if we knew where our students were at.

Two further purposes were identified for NAPLAN. The first related to resource allocation at the state and regional level, using NAPLAN achievement data for comparative purposes to identify areas of need, including low "socio-economic communities" and Indigenous students.

The second additional purpose addressed NAPLAN's role in providing information to parents to show them where their children are at a point in time, in comparison with all "children across the country" who completed NAPLAN at the same time. This comment does raise a question regarding understanding at both educator and parental level of the degree of confidence that can be placed on any individual student outcome at the single point-in-time of NAPLAN completion (Wu, 2016). Participants' perceptions about the role of NAPLAN as a single source of information for individual student diagnostic purposes, while reflecting the original intention of the introduction of literacy and numeracy assessments (MCEETYA, 1989, 1997a), raise the need for caution in current purposes for NAPLAN, although some participants indicated awareness that NAPLAN data cannot "pinpoint" individual student achievement due to the "vast … margin of error". Matters (2018) found that parents did not consider NAPLAN to be a diagnostic test although they did not "fully understand" its purpose (p. 33).

While accountability was viewed by many participants as the primary, and needed, purpose for NAPLAN, it was also viewed negatively, even as "menacing", when considered in terms of uses such as development of league tables and increasing competitiveness rather than collaboration in schooling.

> I think that its purposes are about A: accountability and B: improvement. I think it potentially has, or had, opportunities for the system, the school, and the classroom so, and it's fulfilling its potential which I think has been distorted. It offers those potentials, it is only a small part of the story and the NAPLAN tragedy is that it has been distorted into an organism of its own, a life of its own, and other important things have been lost in that process.

Accountability, the system monitoring purpose, was seen by many to have overtaken "the moral purpose" of students and their learning, creating a high-stakes accountability environment. Similar findings emerge

from previous Australian research with schools, teachers and parents, highlighting "ranking" and "policing" schools and systems as the evolved purpose of NAPLAN (Dulfer et al., 2012; Gable & Lingard, 2015; Matters, 2018; QTU, 2018).

Given the intended range of purposes identified in previous research for systems such as NAPLAN to assist student learning, inform teaching and monitor schools and systems, with possible conflicts between these, School Survey respondents were asked to rank their perceptions of current purposes of NAPLAN and what they thought the ranking should be. The large majority of survey respondents, and organisational representatives, identified current purposes as prioritising system and school monitoring (Figure 4.1). They were equally strong in identifying their preferred purposes as focus on individual students and informing teaching and learning, reflecting the original intention of literacy and numeracy policy to improve learning for all students. These rankings were consistent regardless of the survey respondent's role in school or their school's region or geolocation.

Figure 4.1. Survey respondent rank ordering of purposes of NAPLAN: Current and desirable (row percentages) (n = 5,814)
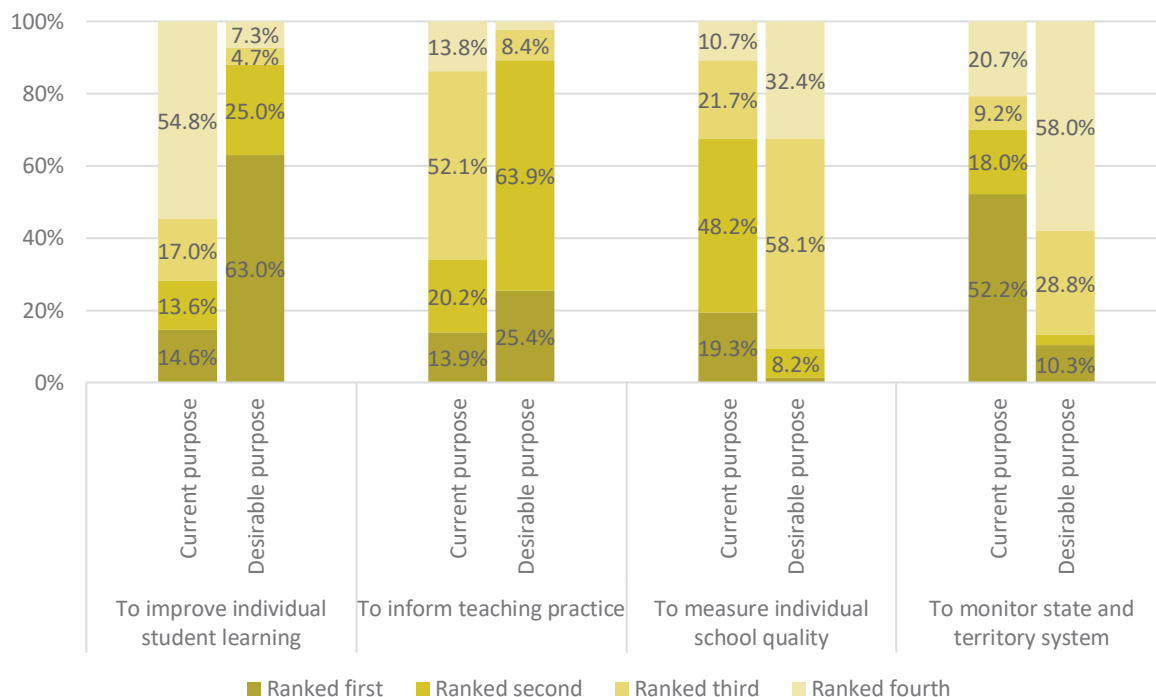
Summary

Consistent with previous policy statements and international and Australian research, participants assigned multiple purposes to NAPLAN, both as a tool and data, including potentially divergent purposes of accountability and informing teaching and learning. Within these two purposes, accountability for system, sector and school level purposes was seen as having risen in priority. Those identifying teaching and learning purposes as priority tended to express their focus more strongly, often referring to the origins of literacy and numeracy testing. However, educational accountability was identified as an expectation of the schooling landscape and of itself not an issue.

Participants identified the range of purposes that NAPLAN data served within systems and schools for improving curriculum programming, teaching and learning. A third purpose identified by participants was to provide parents with information about their children's literacy and numeracy achievement. Provision of information on system and school performance to the wider community was not identified as a major purpose, although discussions of accountability did mention transparency of expenditure of public funding. Several participants expressed a perception that parents appreciated comparative information about their child with respect to their NAPLAN information, although this was not consistent with the findings of Matters (2018).

A caution is noted regarding the NAPLAN purpose identified by some participants for monitoring individual student progress across years of NAPLAN testing. The measurement errors associated with individual NAPLAN scores mean that NAPLAN data alone are not sufficient for consideration of an individual student's level of achievement. NAPLAN tests do not have sufficient reliability at the individual score level to make firm judgements about an individual student's achievement or needs on the basis of NAPLAN data alone.

Value of NAPLAN

In the following section we address the extent to which participants identified the value of NAPLAN to meet different purposes. Term of Reference 4 for Phase 2 of the Review identified as a core issue the "value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level".

Interview and focus group participants identified two clear and general benefits at system level of the introduction of NAPLAN. The most prevalent value was the extent to which NAPLAN had provided the "wake-up call" or "Queensland shock" when Queensland results for the first NAPLAN tests of 2008 were released. References to "coasting" or complacency were made, with follow-up statements that the introduction of NAPLAN made Queensland educators aware that a clearer focus on literacy and numeracy was needed, providing the impetus for change. Further structural changes also led to Queensland school students commencing school at the same age as in other states and hence gaining that additional year of schooling prior to the first year of NAPLAN testing.

> I think it has been beneficial for the state, to be honest, I really do. I think it's been a good thing … I think the first year it came through, that was the first year when we had benchmarks, was when it was like, "Oh, that's interesting" and really then set us on a path for focusing on things like reading.

Clearly, response to the NAPLAN wake-up call served to focus on improving student literacy and numeracy achievement, demonstrated in Queensland's performance from 2008 to present times (ACARA, 2017), identifying a value of NAPLAN data also in monitoring longitudinal trends.

The second major benefit identified from the implementation of NAPLAN was the increased awareness of educators at system and school levels of the value of using data and evidence to inform programs, resource allocation, teaching and learning. Such value has led to increased provision of professional development to improve "data literacy", reflecting a key expectation of the professional standards for teachers (AITSL, 2011). The corollary of this benefit is that NAPLAN has more potential value if more educators have high levels of data

literacy (Carey, Grainger, & Christie, 2017; Datnow & Hubbard, 2016; Pierce & Chick, 2011; Pierce, Chick, & Gordon, 2013). These matters are further explored under *Uses of NAPLAN*.

These two benefits of NAPLAN, both focused on improving student learning, are expressed at a very general level of policy development and implementation. Interview and focus group participants identified more specific values of NAPLAN within their perceptions of its purpose and use. The major values of NAPLAN were related to the availability of "big data", situating literacy and numeracy performance within national and state standards, and regional standards. Value was also seen in the availability of NAPLAN data to assist in teaching and learning programming within schools, identifying gaps and directing resources to where they were needed. Overall usefulness of NAPLAN is discussed under the theme *Uses of NAPLAN*.

Related to the value of NAPLAN and its use to improve learning were discussions by interviews and focus groups of NAPLAN as "one piece of data" that could be integrated with a range of other performance indicators and forms of data. Value in this respect lay in how schools and teachers were able to make use of NAPLAN data. As discussed more fully under *Uses of NAPLAN*, it has been realised in many schools that there is value in using NAPLAN data in conjunction with other sources of evidence about student learning. In this context, NAPLAN is therefore seen as providing value, not in isolation but as contributing to "rich data" that allow creation of narratives about system and school literacy and numeracy performance (Renshaw et al., 2013) and development of an "educative disposition" in some schools (Hardy, 2014a). Value in systems such as NAPLAN is often seen in development of communities of practice around data, reported here as occurring within and across schools by a number of regional focus group participants.

One value identified for NAPLAN was that the move to national testing and national standards of literacy and numeracy also provided the drive for the development of a common national curriculum, the Australian Curriculum. Value was also seen to be enhanced as items within NAPLAN have become increasingly aligned with the Australian Curriculum, achieved from 2016.

A number of Queensland parents in Phase 1 of the NAPLAN Review identified value of NAPLAN for accountability, benchmarking and improving learning, but most did not. Parents appear to have limited understanding of NAPLAN to appreciate its value and benefits (Matters, 2018). On the other hand, some parents identified value in NAPLAN for their own children, through the development of test-taking skills, and implicitly of test-taking resilience, in preparation for later external assessments, especially in high school (Matters, 2018). Previous research has also identified developing such practices as beneficial, as good test-taking preparation involved "planning [and] goal setting" (Thompson, 2013, p. 67).

A considerable proportion of the Queensland parents also considered that NAPLAN had "zero value" for different stakeholders. Those who were seen to gain most from NAPLAN were government systems and schools, with parents, students and the general public gaining little (Matters, 2018). Overall, however, research has shown that parents appreciate gaining information on student achievement (Whitlam Institute, 2013; Wyn et al., 2014).

School Survey respondents were asked to rate the value of NAPLAN (from *no value* to *very high value*) for the range of purposes that have been identified (Figure 4.2). These purposes include provision of support for students in monitoring their own learning, identified in assessment research as a critical element of

effective assessment (Baird et al., 2014), provision of input at class and school level, and for public monitoring of school and system quality. Echoing parent data (Matters, 2018), the weight of opinion of school personnel overall is against the value of NAPLAN. For overall value, 71 per cent said that it had *little* or *no value*, 20 per cent thought it might have *some* value, and less than 10 per cent thought it had *high* or *very high* value. None of the possible uses was seen on average as having *high* value, with averages tending between *little* and *some value*. While the least value was seen in "assisting students in managing their own learning", value for accountability was also seen as limited. The areas where most value was expressed on average related to school uses of the data for programming and teaching, also noted by organisational representatives.

Given the lack of positivity in these responses for NAPLAN's value overall, responses were analysed to see if there were differences across school roles. Consistent with previous research (Dulfer et al., 2012; Wyn, Turnbull & Grimshaw, 2014), although still predominantly negative about the value of NAPLAN overall, school principals and deputy principals were more positive in rating NAPLAN as having *some* to *high value* (48.6%) than heads of school or curriculum (41.1%) or teachers (24.2%). *High* and *very high* value was seen by one in six school leaders versus one in 13 for other staff.

School Survey respondents made several comments that indicate what may have affected their perceptions of the value of NAPLAN and may have contributed to the negative perspectives reported above. They commented on the limited availability of time to adequately analyse NAPLAN data when results were received. Also, the time lag between testing and results was noted to reduce the value of NAPLAN for school programming since the majority of students had progressed in their learning. The time lag factor has been noted to affect teacher engagement with NAPLAN, even when expectations for the potential value of NAPLAN data are high (Pierce & Chick, 2011).

Figure 4.2. Perceived value of NAPLAN for a range of uses (row percentages) (n = 5,814)

	No value	Little value	Some value	High value	Very high value
Assisting students in managing their own learning	55.2%	31.9%	10.1%		
Showing students how well they are progressing	25.6%	35.4%	30.4%	6.4%	
Showing teachers where to inform their teaching practice	16.2%	27.7%	40.1%	11.8%	
Helping teachers identify individual student learning needs	21.3%	32.9%	31.7%	10.4%	
Providing input at the school level to discussions about program improvement	17.7%	29.3%	35.5%	13.0%	
Identifying at the school level areas of learning where more attention is needed	12.1%	23.2%	40.6%	17.7%	
Showing the community how well the school is teaching the students	46.9%	30.6%	15.9%	4.4%	
Informing parents about their child	24.7%	36.2%	29.2%	7.0%	
Holding schools accountable for student learning	43.6%	29.9%	17.7%	5.4%	
Holding principals accountable for student learning	43.5%	30.5%	17.3%	5.5%	
Overall value	30.9%	40.2%	20.4%	6.2%	

Many students who completed the Student Survey provided comments related to the value of NAPLAN. Almost all were negative, with almost a quarter of comments indicating that it was "a waste of time", and that they "didn't try" as it was not on their report card. Students echoed School Survey comments that teachers had information through their school assessments as to what they knew and could do, and their strengths and weaknesses, as did parents. Some indicated that such assessment information should be sufficient for essentially accountability purposes, rather than a single test that was stressful for them. Students considered that NAPLAN did not help their learning: "From my personal experience with the exams, I have gained little-to-no educational value".

Value of NAPLAN and NAPLAN validity

Several School Survey comments regarding the value of NAPLAN reflected concerns with what could be considered validity. Most respondents described NAPLAN as an inaccurate representation of the school, the teachers, and student ability. The design and timing of testing were two factors that contributed to this inaccuracy. The multiple-choice questions were described as providing inaccurate results—some

respondents described students guessing if they did not know an answer or if they were running out of time. The design of the writing criteria that privileged complexity in vocabulary was considered as not showing the true skills of students to construct engaging and creative texts. NAPLAN was described by many participants as "a time test", not a measure of student capability. Some considered NAPLAN to be culturally biased and suitable only "if you are a student" who is a "second generation, native English-speaking Australian of middling to upper socio-economic status, who lives in a major city, has no disabilities, are neuro-typical, and have had no incidents or illnesses that have caused a disruption to your learning". NAPLAN was also described as not accommodating or taking account of the differential development of skills in young children—"developmentally inappropriate to expect all students at this age to have a specific set of skills"—which was described as contrary to learning and child development theories. The tests were described as not reflecting a student's differing abilities, nor the contribution of a school to a student's learning. The number of variables that can impact results was frequently mentioned. This included students who were uninterested in performing well. Seventy-nine per cent of students indicated that they did not think NAPLAN outcomes were a true reflection of their ability.

NAPLAN was viewed by respondents as failing to measure the skills valued in current times, including the skills to "think deeply or creatively". Literacy tests were described as morphing over time, with increasing amounts of texts that failed to take into consideration the fatigue of young children. In particular, the Reading test requirements in terms of the length of texts and questions were criticised as too long for Year 3 students. The time limit on the tests was described as a hindrance for all students, but specifically for children who have learning disabilities and who, with greater time, could demonstrate their knowledge and skills. Some noted that the educational adjustments available in classrooms for students with disability cannot be provided in NAPLAN testing, with its strong focus on consistency of administration. However, some respondents also noted that the extra time provisions were ineffective for some students, for example those with dyslexia who required "the option to access the reading test aurally and the option to respond orally".

Some participants noted that catering to individual learning needs, as typically undertaken in everyday classroom practice, was not a feature of NAPLAN testing, rendering the test invalid for many respondents. Significantly, inadequate time for the Writing task was highlighted by the majority of respondents who identified that the otherwise valued aspects of the brainstorming and discussion, planning, drafting and editing processes in which good ideas and good writing surface could not be reproduced within NAPLAN. The formulaic approach to writing and the limited number of text types (in particular, persuasive writing) were further criticisms, with many respondents describing NAPLAN writing as diminishing the quality of student writing in curriculum. Many respondents called for more authentic writing tasks and text types relevant to the age group. For example, some respondents suggested that "recount" would be a more appropriate writing task for Year 3 students. We address NAPLAN and Writing more fully in a later section.

Value of NAPLAN: Experience and assessment identity

How school personnel value NAPLAN may also be affected by their reported NAPLAN experiences. School Survey respondents reported on the quality of their first and most recent experiences with NAPLAN. Table 4.1 shows that both experiences were generally unfavourable: 44 per cent of respondents said that their first

experience of NAPLAN had been negative and only 19 per cent recorded a positive first experience. A particularly striking feature of experience with NAPLAN is that respondents' recent experience was worse than their first experience: 57 per cent said that their recent experience of NAPLAN had been negative and only 15 per cent said it was positive. Perhaps not surprisingly, the Overall Value School Survey respondents assigned to NAPLAN was well-correlated with the negativity or positivity of their experience, that is, the more positive the experience, the higher the value noted. This was most notable for their most recent experience.

Table 4.1. School Survey respondents' first and most recent experience of NAPLAN (column percentages)

	First experience	Most recent experience
Very negative	15.4	22.4
Somewhat negative	28.1	34.7
Neutral	38.0	27.5
Somewhat positive	15.0	10.7
Very positive	3.5	4.6
Total responses	5,812	5,805

An interesting outcome of the perceptions provided through interviews, focus groups and surveys was the individual nature of educators' response to NAPLAN. As the analysis of the range of views present within a sample of schools showed, views regarding the value of NAPLAN varied within roles within schools as well as across roles and schools. Recent research on teachers and assessment has focused on teacher assessment identity (Looney, Cumming, van der Kleij, & Harris, 2017), establishing that how teachers approach assessment practices and innovation reflects not only their assessment literacy, in terms of knowledge and skills, but also their confidence, their own experiences of assessment and emotional responses to assessment.

Value of My School

Interview and focus group participants made a number of comments that referred to the value of My School for NAPLAN, including looking at trends over time, and identifying schools that were doing well. While most noted limited use by parents, others noted that some parents seemed to find value in the comparative performance of their child's school with others. School Survey respondents were asked to identify the usefulness, or value, of My School for school staff (Figure 4.3) and for parents (Figure 4.4). While the majority perceived NAPLAN to have limited value for schools, approximately 30 per cent of respondents felt My School results had some use for schools, with fewer considering My School of value for parents.

Figure 4.3. Usefulness of aspects of My School for principal and teaching staff in your school (row percentages) (n = 5,814)
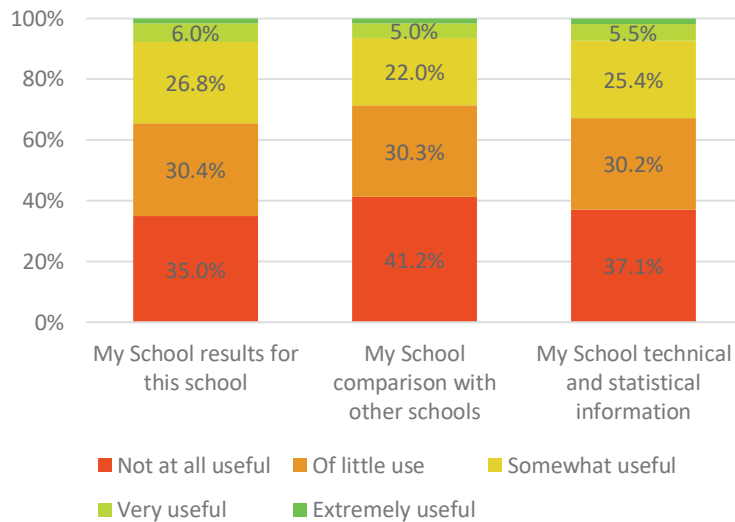
Figure 4.4. Usefulness of aspects of My School for parents in your school (row percentages) (n = 5,814)
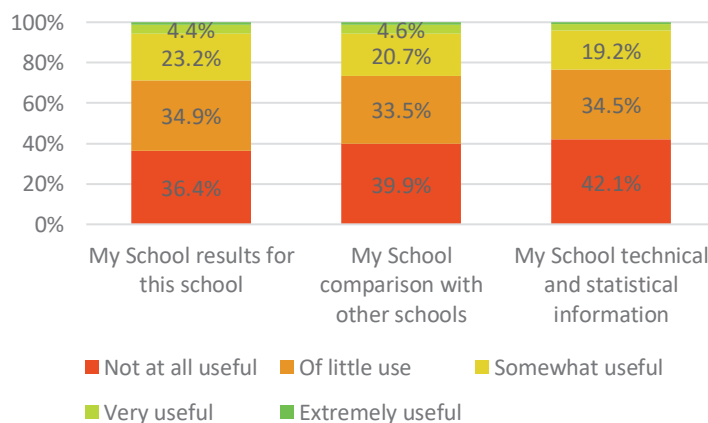
Use of NAPLAN data

Development of a culture of data use

At the system level, it is generally thought that NAPLAN data have been successfully *used to introduce reforms*, resulting in improved outcomes. Initiatives that are thought to have mattered include curriculum initiatives, reading program initiatives, and school improvement planning. Even so, there were some critical comments about the top-down nature of such reforms, the narrowness of the literacy and numeracy focus, and the over-dependency on a "point-in-time" or "snap-shot" of "some aspects of curriculum". Others value NAPLAN as at least giving some information about comparative performance and standards that allows diagnosis of areas for improvement—"a starting point, even if it is not everything". The School Improvement Unit of the Department Education uses NAPLAN as only one of a "myriad of data sets" to determine school need for resources, aiming to make NAPLAN less high stakes.

In interviews and focus groups it was thought that there has been *a change in the discourse about NAPLAN*, deliberately moving to be more constructive in outlook, with more encouragement to reduce the focus on NAPLAN and to focus instead on quality curriculum and quality teaching. There may, though, be some distance to travel before this is endemic; this renewed focus has not yet permeated all schools and it would seem that NAPLAN is still the implicit goal, with much talk remaining, particularly at school level, about meeting targets and lifting NAPLAN performance. It would seem that there is probably a four-year lag in take-up of this change and that policy may not yet be well articulated and communicated.

NAPLAN results are still seen as being key performance indicators. While the discourse at the higher levels of education sectors was that NAPLAN was only one indicator of school performance, principals and aspiring principals were noted as needing to demonstrate NAPLAN improvements as part of their performance management plans. This also applied to higher levels of management. Similarly, most schools were considered still to need to include NAPLAN improvement targets in improvement plans.

More positively, as noted in *Value of NAPLAN,* a major perspective of the interview and focus groups was that the introduction of NAPLAN has led to *greater awareness of the use of data* to inform system and school planning and classroom teaching. This began with NAPLAN data but spread to include other forms of data. A range of uses of NAPLAN data in conjunction with collection and interpretation of other data, both assessment data and contextual data, was claimed to be occurring widely.

Nearly all school-level participants and many management level participants commented on the use of NAPLAN to "triangulate" with other data, including different forms of assessment data from classroom assessments by teachers and levels of achievement (LoA), to other external tests or processes that schools reported using (e.g. Progressive Achievement Tests (PAT-M, PAT-R), and PM Benchmarking). Multiple references were made to using PAT-M and PAT-R testing and data. NAPLAN data formed part of data profiles from different sector-level statistical resources. Focus group participants referred to "vast arrays of data sets" in schools.

> … from a school perspective the purpose is to triangulate that data with our LOA data. And I guess there's some discrepancies sometimes around whether that does triangulate or not.

> So, all of our schools collect PM benchmarks from Prep to Year 2. We do writing analysis from 3-10, so that's four captures a year that are co-moderated and marked against eight criteria which align pretty closely with the NAPLAN criteria. PAT-R, PAT-M pretty much and we're doing numeracy monitoring will come on board soon.

The focus groups reported engagement with writing data and how the writing data was utilised as a resource to inform next-step programming and teaching. The participant below was cognisant of the need for improvement in writing and indicates the need to "drill down" to components of the NAPLAN writing test such as the "marking guide" as a resource for learning,

> I think that schools generally appreciate too the ability for NAPLAN to drill down, for example, if the improvement agenda is around writing, the NAPLAN marking guide and what NAPLAN provides

for schools to be able to drill down again to determine whether what they're doing is working well or not, I think schools appreciate that.

With greater use of data to inform decision making comes a *need for increased capability in the use of data*, for both school leaders and teachers. It is perceived by some that NAPLAN data have not been well used and often misused. While there were statements that there has been some action on building capability, there is also a felt need for more deliberate and coordinated attention to this. Particular mention was made of the need to build data literacy skills for interpreting and using data. One person noted that other countries are moving to build data literacy into teacher training and professional development. The need, in building an evidence-based culture, for school leaders and teachers to acquire data literacy skills was noted in the research literature, together with the complexities that are involved in doing so (Carey, Grainger, & Christie, 2017; Datnow & Hubbard, 2016; Pierce and Chick, 2011).

While various sources of literature are still indicating that there is a need to improve school leader and teacher data awareness, a substantial number of respondents to the School Survey claimed that they understood NAPLAN data either fairly or very well, with school leaders more confident than teachers (Figure 4.5).

Figure 4.5. Understanding NAPLAN data for role in school and overall (row percentages)
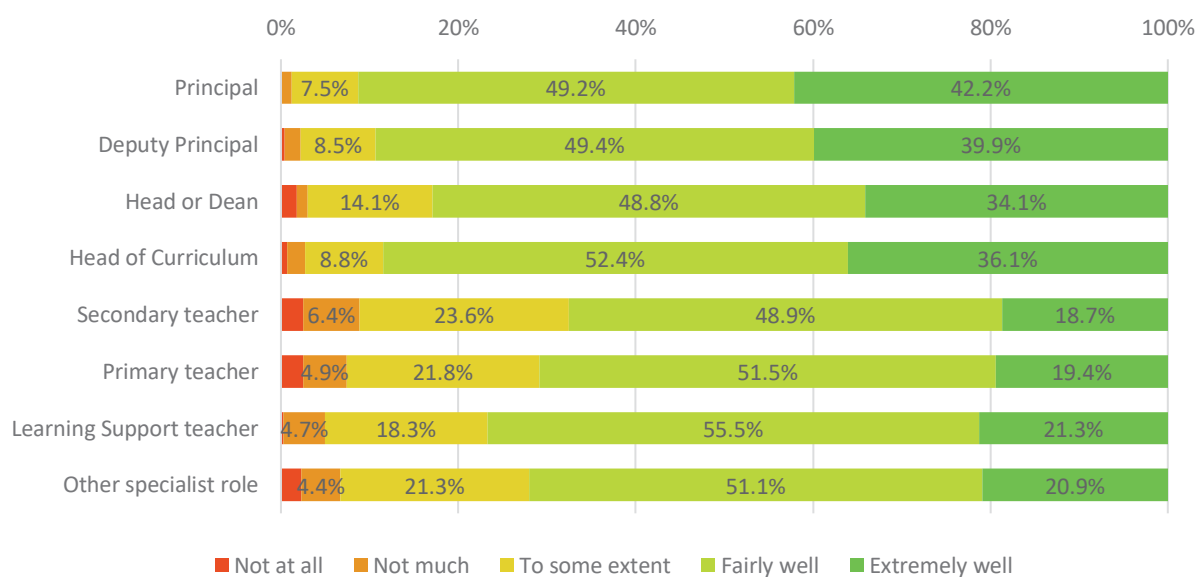
Confidence among teachers in understanding of NAPLAN data, not surprisingly, varied with the number of years of teaching experience, beginning teachers being less confident than experienced teachers (Figure 4.6). This indicates much more confidence in understanding NAPLAN data than the research on data literacy has previously identified among teachers (Pierce & Chick, 2011). This does not diminish the need for more professional development in data literacy.

Figure 4.6. Understanding NAPLAN by years of teaching experience (row percentages)
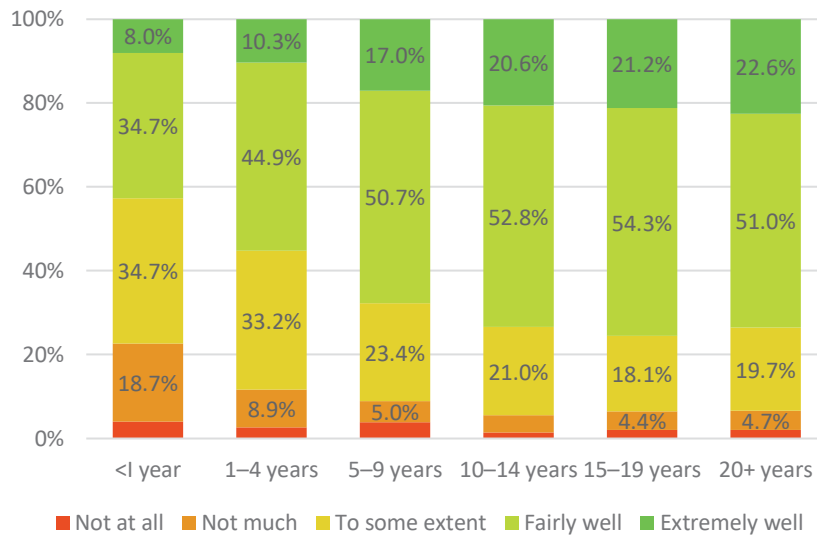
Within-school engagement with NAPLAN data

In the School Survey, school leaders reported a substantial amount of activity with NAPLAN data in their school, including their own direct involvement (Figure 4.7). The general expectation was that teachers would be involved in interpreting and analysing NAPLAN data, and doing so collaboratively. There was also an expectation that this would lead to various actions. There was much less expectation of using external specialists to help in such matters. Also, there was much less intention to engage the school community (parents).

There was a lack of match between these expectations of school leaders and the practices reported by teachers. Teachers indicated somewhat less engagement with NAPLAN data than school leaders expected, on average to somewhere between *not much* and to *some extent* (Figure 4.8). More than one-fifth of the teachers were involved in each of these activities a *fair* or *substantial* amount. In fact, more than one-quarter were involved in four of these activities a fair or substantial amount: interpreting and using NAPLAN data; analysing individual student strengths and weaknesses; analysing class strengths and weaknesses; and changing teaching strategies to improve performance. However, at the other end of the scale, somewhat similar numbers of respondents were not at all involved in any of these activities. Overall, this indicates limited use of NAPLAN data by teachers. However, there is a spread of practice right across the spectrum.

The interviews and focus groups provide an elaboration of these School Survey results, also indicating that within-school management approaches to NAPLAN data use can vary substantially. There is, however, widespread recognition that successful data use requires *school leadership, focus on learning, and teacher collaboration* (as identified in the research literature). Principal leadership was recognised as critical for data use, setting the tone and expectations, such as a learning culture. One principal said:

> … it's our job to actually facilitate that dialogue to encourage, to coach, to mentor, to ensure that that happens.

Figure 4.7. School leader expectations of engagement with NAPLAN (row percentages) (n = 1,311)
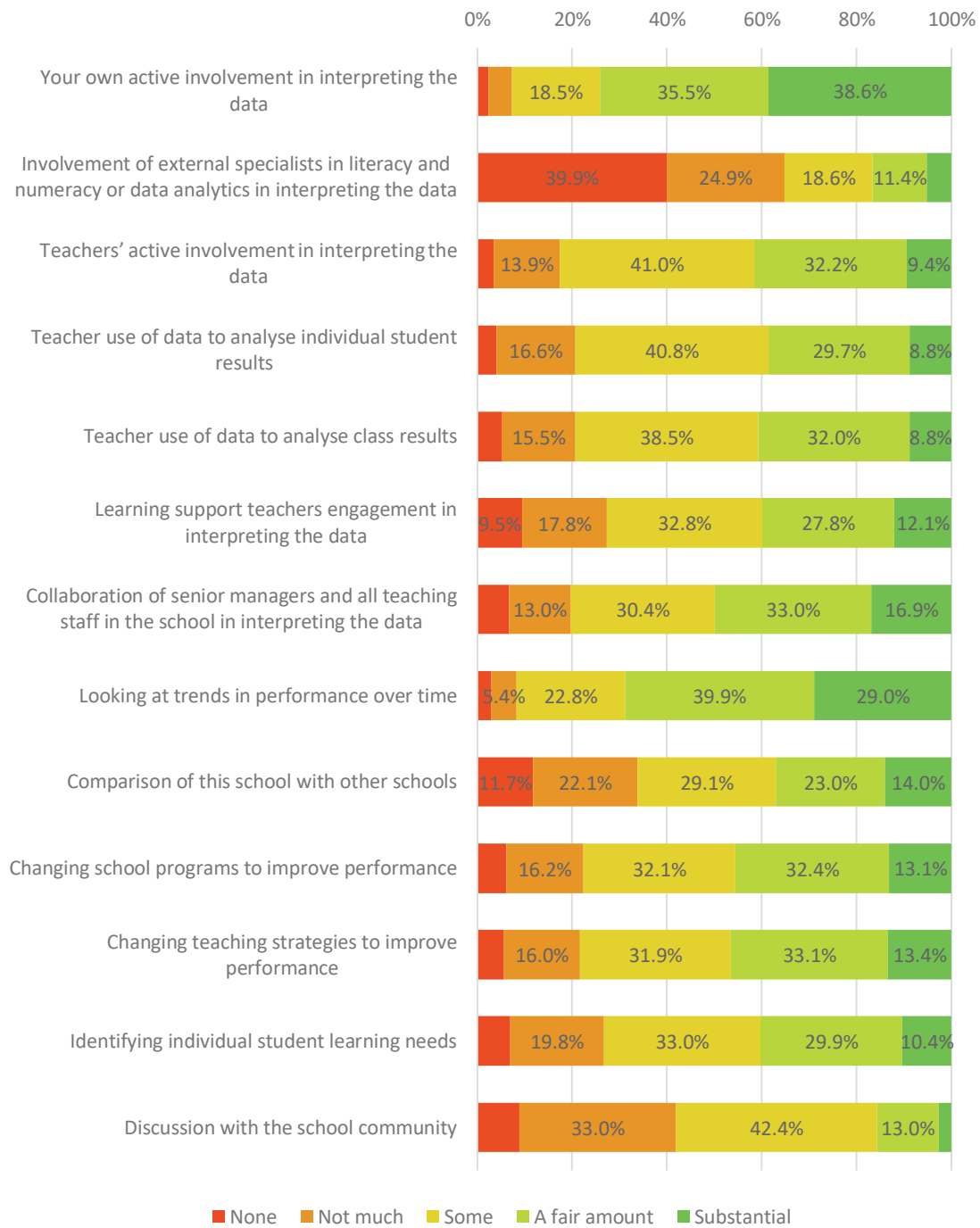
Figure 4.8. Teacher engagement with NAPLAN (row percentages) (n = 4,503)

	None	Not much	Some	A fair amount	Substantial
Interpreting and using the NAPLAN data	14.9%	20.8%	35.0%	21.6%	7.7%
Accessing NAPLAN data files for your own analysis of the data	21.1%	20.4%	30.4%	19.9%	8.2%
Analysing individual student strengths and weaknesses	18.1%	21.3%	32.4%	20.8%	7.4%
Analysing class strengths and weaknesses	19.3%	21.9%	30.8%	20.6%	7.5%
Analysing class trends in performance over time	23.7%	25.6%	28.4%	16.2%	
Collaborating with senior management in interpreting the data	27.3%	23.1%	27.3%	16.0%	
Collaborating with other teachers in data interpretation	23.6%	24.8%	29.8%	16.7%	
Collaborating with other teachers in acting on the data	25.2%	24.9%	28.6%	16.5%	
Changing teaching strategies to improve performance	20.7%	21.5%	32.8%	18.8%	6.4%
Using the data to address individual student learning needs	23.5%	24.0%	30.1%	16.9%	

Some principals became engaged in data analysis but preferred to keep teachers at distance from the data. In these cases, NAPLAN data were "structured" and "scaffolded" as "useful pieces of information" before provision to teachers. While this is not generally considered in the research literature as the best strategy for engaging teachers with the data, one principal said:

> … the real strength is the capacity of the principal to be able to interpret the data, deliver it to the staff, … engage staff in that process so that they're able to analyse the data effectively to make the difference in the teaching and learning area for individual students.

Institute for Learning Sciences and Teacher Education 2018

The focus groups identified different ways in which teachers were engaged in data use. Practice was said to vary between teachers from only in certain years to all teachers being engaged, in that case with data use promoted as "everyone's business" and about "getting everyone on board". One principal said:

> I do find a sense of staff ownership beyond the two focus year levels … It's part of a celebration of the hard yards they've done but also recognising the results of Year 3 do begin from prep and I know that the [Year 4s] feel a huge ownership of the "What's happened? Who's come through? What are the implications for Year 5?" So, I think it's a really big collective moment, again, at my school.

There was also a view that there are different degrees of staff "buy-in" and that this can depend on the school context, with some schools not giving NAPLAN data use priority, others only attending to whole school data, others again having only middle leaders involved, or in secondary schools only English and Mathematics teachers. Some concern was expressed about the pressures of time preventing engagement. In primary schools it may be difficult to engage the interest of Preparatory and Year 6 teachers.

Collaboration among teachers was seen as encouraged in some schools and as a known effective practice for data use. This is sometimes enacted through a "professional learning team" approach, where there is an emphasis on making collective use of data to improve practice. However, teacher participation in collaborative professional learning teams was not indicated to be standard practice. Some comment was made about the use at system level of an enquiry cycle for data use, with NAPLAN data just one of several types of data in a school profile. In this case, the focus was on school improvement, but the enquiry cycle is also relevant for teachers within schools, implementation more broadly was not much in evidence.

Collaboration has been more prevalent at the inter-school level, even though it is noted that competition among schools can work against this strategy. Small schools have developed collaborative working relationships to share and develop resources, especially through coaching and mentoring. Details of such processes are identified in existing research (Datnow & Wohlsetter, 2007; Earl & Katz, 2002; Cumming, Maxwell, & Wyatt-Smith, 2016; Maxwell, in preparation; O'Day, 2002; Sutherland, 2004; Wahlstrom et al., 2010).

Within-school interpretation and use of NAPLAN data

As a follow-up to the question on engagement, the School Survey asked specifically what use individual teachers made of NAPLAN data. The results erode the indications of engagement with data previously indicated in Figure 4.8. A large proportion of respondents did not use NAPLAN data to any extent for most of the possibilities listed, with somewhere between *none* and *not much* being the most typical average response (Figure 4.9). The largest statistic is that about three-fifths of teachers were very disinclined to use NAPLAN data in preference to their own classroom data—not unexpected in light of previous research (Cumming, Wyatt-Smith & Colbert, 2016; Cumming et al., 2006; Hardy, 2013). This is supported by a substantial minority of teachers (about two-fifths) seeing NAPLAN data as not relevant for confirming their own judgements. NAPLAN data were seen as providing very little new information and even the band categories received little attention. On balance, responses to the question indicate that teachers do not see NAPLAN data as very relevant, informative or useful.

There was more evidence of engagement with data on one item—use of other diagnostic tests—indicating some take-up, confirming other findings (Hardy, 2014a, 2015a, 2015b, 2018; Hardy & Lewis, 2017b; Lewis & Hardy, 2016; Lingard, 2010; Klenowski & Wyatt-Smith, 2012; Ragusa & Bousfield, 2017; Singh et al., 2015) about widespread use of such tests, However, this engagement is equivocal (almost as many on the negative side as on the positive side).

Figure 4.9: Teacher interpretation and use of NAPLAN test data (row percentages) (n=4503)
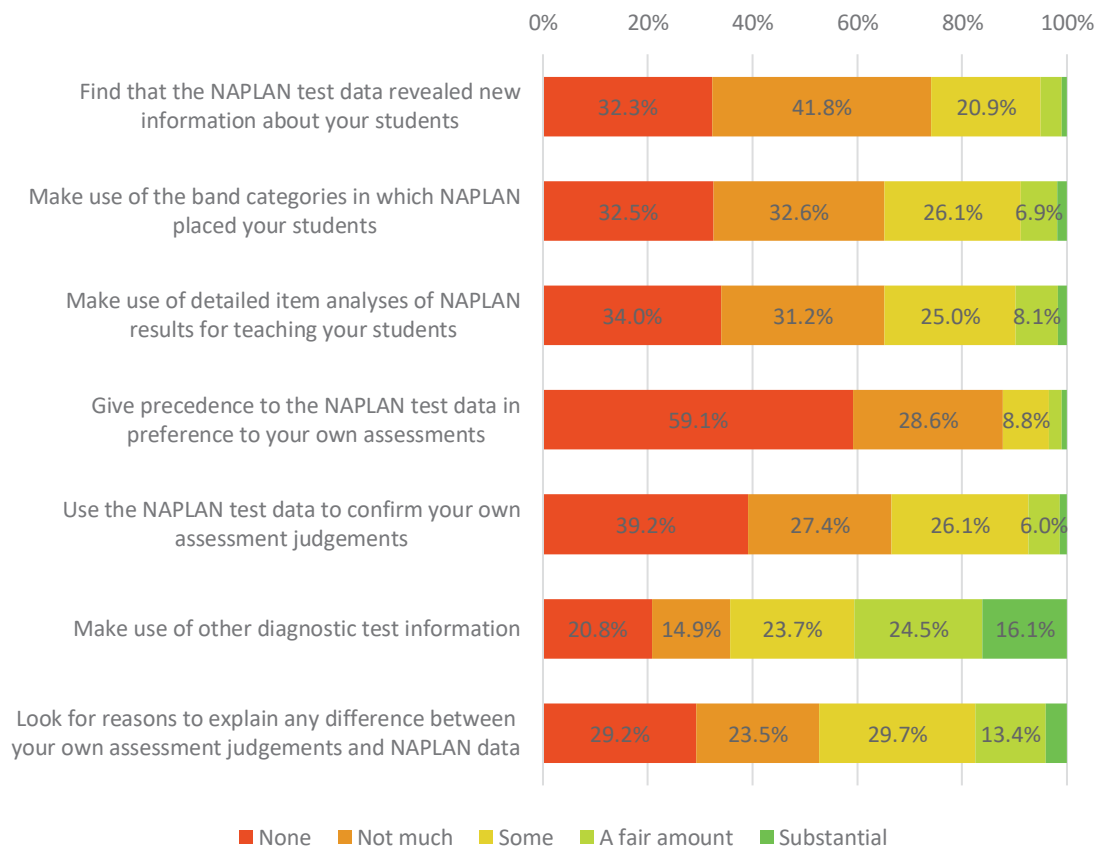
The interviews and focus groups provide some elaboration of these School Survey findings, also indicating a variety of perspectives on how NAPLAN data are being used. In relation to the issue of *whether NAPLAN data or school assessments should have primacy*, some feel that NAPLAN data have primacy, trust the test designers to produce "accurate" results, and focus on the implications of NAPLAN results for educational practice. NAPLAN results are also used for the "validation" of classroom assessments, drawing confidence in these assessments when there is a "high correlation" with NAPLAN, especially in relation to high achieving students. Others have an opposite view, considering NAPLAN to be of limited value for showing what students can do, at best confirming what is already known, and preferencing their own judgement where there is any contradiction (found also to be prevalent in other accountability systems, see Chapter 2). From the School Survey (see above), privileging classroom assessments appears to be dominant practice.

It is reported also that a common approach is to use NAPLAN data in conjunction with other data. This approach is supported by the view that NAPLAN provides just one data set, though an important one,

among a whole range of data sets that schools collect. Schools are seen as being "awash in data", with "a vast array of data sets to help us make decisions", leading to a reduction in use, or at least a repositioning, of NAPLAN. The concern is that there may be too much data (one reference was to "bowerbirds" and "not being sure what to do with all the data"). Clearly, some of it, especially the data obtained from commercially available tests, may lack validity in not being linked to the Australian Curriculum or to learning progressions. Some schools are being more selective than others, "focusing on particular data rather than just a full suite of everything". However, there is also an apparent tendency to collect and use such data on short (three to six week) cycles, rather than taking the long view. The overall concentration on additional test data, in preference to more authentic forms of assessment demonstrates the performativity influence that NAPLAN has had. While the focus of Phase 2 of the NAPLAN Review was on NAPLAN, potentially affecting the range of discussions, the lack of discussion around improvements to the quality of school assessments also suggests that, beyond professional development in data literacy, there may be continuing need for professional development in assessment literacy, to build skill and confidence in assessment.

Some schools, however, are reported as taking a more direct approach to using NAPLAN data in conjunction with other data. A very common reference is to data triangulation, a key concept noted in previous research (Renshaw et al., 2013). In some cases, the term is being used inappropriately—taking it to mean checking whether two measures align (validation); this may indicate further need for professional development on this point. Others indicated that triangulation involved uses of different data in combination, each making a contribution. In some schools, combined datasets are being used in this way to identify individual student learning needs and to inform teaching directions within schools and classes. There is also talk about the tremendous conversations or "rich dialogue" generated by comparing different data sets and examining their meaning in terms of the different information they provide. Two representative and articulate comments about use of data triangulation were:

> … we use a lot of checking for understanding where we have unpacked the achievement standards with the content descriptors and have very clear guides to making judgements across all of our curriculum areas that are linked, can be backwards mapped. So, a lot of our teachers engage in that constant monitoring of student learning.

> … also using the literacy continuum as a way of moving students throughout that continuum and identifying specific goals and benchmarks as they need to move in literacy across the school.

Many talk about the way data need to be contextualised within the school setting, looking for the story behind the data. From this perspective, NAPLAN is seen as adding "another slice of data … to add to teacher classroom assessment and stories of a child's progress". One expression was "developing a culture of evidence gathering for each child". The relative preference for classroom assessment data rises where NAPLAN is seen as "narrow in focus" and where there is confidence in teacher judgements. Some change away from using NAPLAN as a "benchmark" is recognised as having occurred in recent years.

NAPLAN data are seen to be of more direct use at classroom and individual levels for reflection on areas where more improvement may be needed, as well as celebration within the school of what has gone well (but often with limited action taken as a consequence). Some mention "drilling down" to uncover detail in

the data. This can include analysing individual student performance at the item level, cohort performance at the item level (across classes and years), and tracking students across years (even longitudinally across Year 3 to Year 9). Some comments tie this back to the originally intended focus of NAPLAN on individual student diagnosis. For example:

> … being able to identify where a student, individual student, is at so that that would inform the teacher about practice to get that student up to a reasonable and necessary standard for effective learning to occur … I think that the data in, itself, for an individual student is powerful …

Some, however, recognise that data at these levels are highly unreliable, and that basing firm conclusions on such data can be hazardous. One focus group recognised unreliability (producing variation from year to year) as a problem if NAPLAN is used to hold small schools accountable for *outcomes*. However, sometimes this is ignored, especially when using NAPLAN data to track individual or cohort growth. Others are aware of the "margin of error" for individual students, claiming that NAPLAN was never intended for that. One comment was:

> [as a measurement instrument] it's not a diagnostic for an individual student … it cannot be used to assess the individual performance over time … in terms of diagnosing any strengths or weaknesses, it is not strong enough or robust enough …to be able to do that.

A similar concern was expressed that NAPLAN data are being used in some schools to evaluate teachers or teaching. This is clearly invalid—in most schools, teachers of students in the NAPLAN testing Year levels will have had limited engagement with those students; student performance has many influences, including the work of many teachers; and the class measures have large "margins of error". While some hold a view that NAPLAN can be used to identify better or worse class performance, and want to ask what can be learned about teaching that can be replicated from the better classroom across other classrooms, this is unlikely to be a successful strategy. Some thought that it might be possible to be more strategic in such situations by first seeking to explain the data. Others thought it better to use the data collectively.

There is much talk about classroom assessment—the "shorthand" (as identified by one participant) reference is "A to E". This is bolstered by some reference to curriculum. One expression of this is that since NAPLAN aligns with the curriculum, an emphasis on teaching the curriculum well should lead to improvement in NAPLAN results. However, such talk turns quickly to an emphasis on "the As and Bs" and their relation to the "upper two bands" on NAPLAN: "Show me how you're lifting your As and Bs in Science and that will be a better indicator that you'll get a great upper two bands". There seems to be a strong focus on getting students into the upper two bands: "Mean scale score and the upper two bands are what we really concentrate on". Some identify value in asking why an A or B student is not in the upper two bands: "Are our assessment standards wrong or did the student have a bad day?" There seems to be some focus on both getting C students up to A or B, and getting NAPLAN middle band students into the upper bands. There is little reference in the discourse to D and E students and to NAPLAN lower bands, although one participant noted "[and] those from the lower bands have come into the middle band". Conversely, one participant noted that "[there was] four per cent of kids sitting below national minimum standard. And that was unlikely to ever change, yet our investment, our whole strategy for the first six, seven years of this

had been focused on that." The overall emphasis appears to be on "high achievers" rather than "low achievers". Students of difference are largely absent from the discourse.

Another aspect of the shift towards curriculum-based assessment is still the concern that teacher judgements may be flawed, and that this necessitates the use of an external "validator" such as NAPLAN. A view was expressed, however, that the missing ingredient is systematic moderation processes. One administrator said:

> [The question] is how do we have a quality assured process in our A to E? You know, we're moving there. We're moving a lot better. We're certainly seeing a greater number of schools undertaking moderation, shared moderation across classes, but there's other opportunities to make that more supportive across the system without the unintended consequence of making it more difficult at the classroom level or difficult or high stakes for the teacher.

Another perspective offered is that the emphasis needs to shift from NAPLAN to quality curriculum and teaching. But this requires not only that teachers be able to interpret data but also that they be able to use it to make more effective instructional decisions. One administrator put it this way:

> … the biggest step in NAPLAN going forward is about moving from the "what's" to the "how's" and actually having coordinated and targeted professional learning. So, if I'm a teacher and I have a set of results come back to me that tell me that a lot of kids can't do a particular thing, well I've got all the "what's". But if I don't have the wherewithal in the "how" about what to do in the classroom next week and next month to do that, … this is about making sure teachers actually have the range of strategies necessary to teach the things that they have to teach the kids. And I think the greatest issue in NAPLAN at the moment is to get away from the bag of "what's" and get into the "how's".

Communication of NAPLAN data

School Survey respondents were asked about their engagement in communications of NAPLAN results (Figure 4.10). There are clear differences between school leaders and teachers. Predominantly, both school leaders and teachers attach little to no importance to any of these forms of communication. Perhaps surprisingly, teachers attach even less importance to class and individual discussions about NAPLAN their results. School leaders attach more importance to communications with parents, but mostly about general matters—what NAPLAN tests and do and do not test, as well as the school's response to the results. There was very little importance attached to comparisons with other schools

Figure 4.10. Importance attached to different forms of communication by leaders and teachers (row percentages) (Leaders n = 1,311; teachers n = 4,503)

No importance ■ Little importance ■ Some importance ■ High importance ■ Very high importance
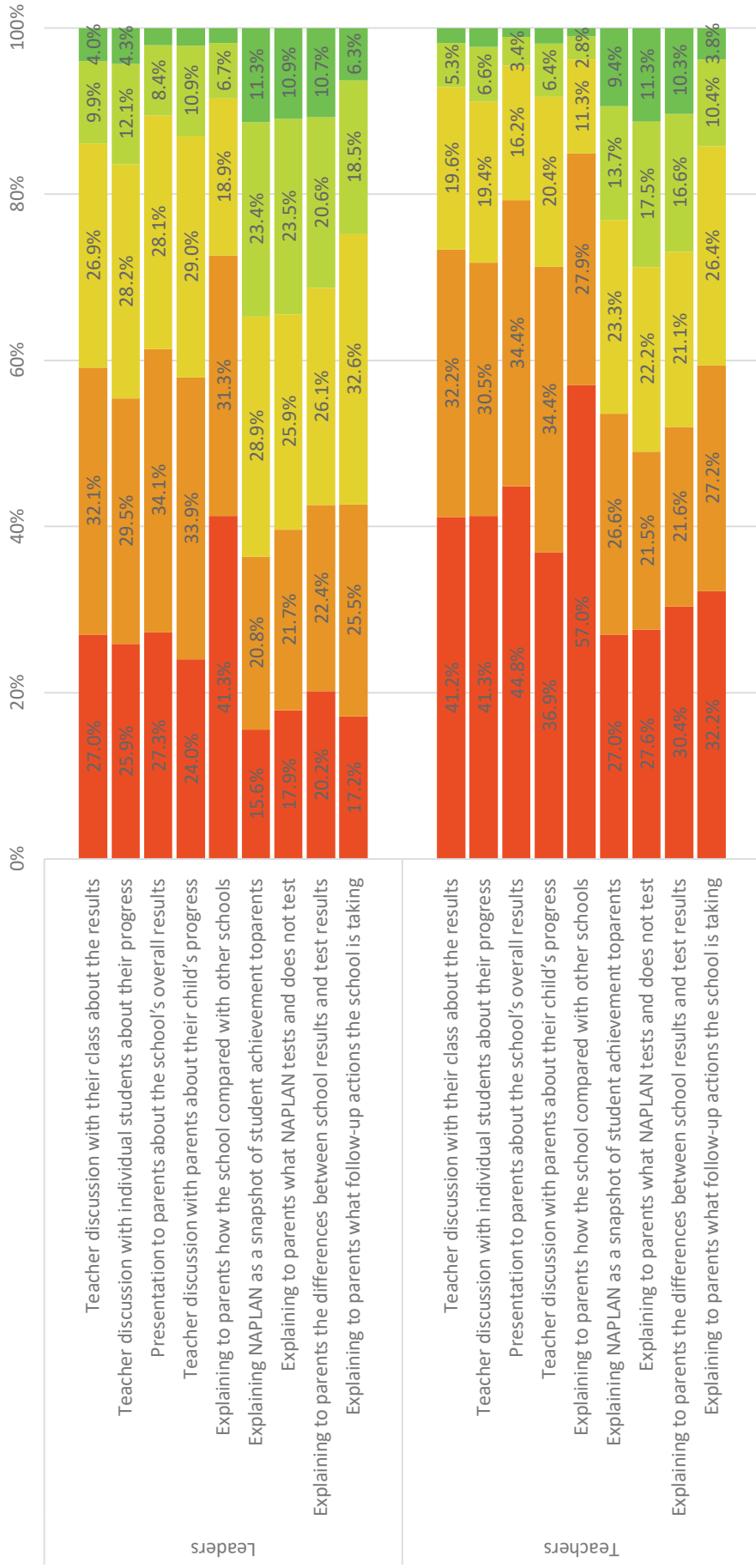
Organisation representatives were more divided about the importance of communication between teachers and students but thought it more important for teachers to explain to the community what NAPLAN measured and differences between school and NAPLAN results.

Other issues in data use

National Minimum Standards: These are considered as no longer useful. For example:

> … sadly, national minimum standard still sits there, and you'll read it in the Queensland reports, because it is the standard where Queensland looks like it's making progress. I personally would question how useful the national minimum standard is because it is too low to be of any value to any student.

My School: Most thought in general that My School was not of much use.

Time lag for results: One of the consistently reported barriers to appropriate use of NAPLAN data has been the time lag between testing and results, making the data to a large extent obsolete. Computerised testing has the capacity to make data more relevant and useable.

Media use of data: Interview and focus group participants noted, with strong fervour, what they considered to be misuse of NAPLAN results by the media.

Summary

It is thought that NAPLAN is more often being used in more sensitive way than previously, as one piece of evidence among many. It has been suggested that there is an on-going shift in the discourse about NAPLAN to reduce the high-stakes importance of NAPLAN and adopt a more constructive approach to the use of data. This does not seem to have progressed far, since NAPLAN is still in widespread use as a key performance indicator; practice may lag behind official policy change by several years. Movement towards a more sophisticated culture of data use in schools requires attention to professional development in data literacy and data use. School leaders and teachers appear to be more confident about their data skills than previous research would indicate.

School leaders report that they generally expect considerable activity in their school in relation to the NAPLAN data, including their own active involvement and collaboration among teachers. There was encouragement for the data to be used to some extent to improve teaching and learning. Application at school or program level is more common than at classroom or student level. In practice, the teachers say that they are much less engaged with NAPLAN data than school leaders expect and generally fairly modestly, though there is a strong minority who say they are more engaged. On the other hand, there is much less engagement with detailed uses of the data; they would seem to try, but find the data wanting. There was a strong rejection of the relevance and usefulness of NAPLAN data (since it is not perceived as telling teachers much they do not already know), with a preference for within-school classroom assessments.

School leadership practices in encouraging data use vary considerably, from close guarding of the data to collaborative involvement. There is some recognition of the importance of engaged leadership, an inquiry culture, and collaborative endeavour, as identified for successful data use in the international research

literature (Cumming et al., 2016; Katz & Earl, 2002; Maxwell, in preparation). Small regional schools, in particular, report inter-school collaboration for developing and sharing resources and capabilities, as is reported in some research studies reported in Chapter 2.

A variety of perspectives on use of NAPLAN data have been identified. There would seem to be a trend towards using NAPLAN data in conjunction with other data—identifying individual student learning needs and informing teaching directions within schools and classes by comparing different data sets and examining their meaning in terms of the different information they provide (a process of "triangulation", as recommended in international research, see Chapter 2). Some practices are, however, concerning. These include: using NAPLAN mainly to "validate" classroom assessments rather than as assessing only some aspects of the curriculum; potentially using too many standardised tests that lack validity in relation to the curriculum; concentration on turning moderate achievers into high achievers with limited attention noted for low achievers and student of difference; failing to recognise the "margin of error" of NAPLAN data for individual students, small groups and schools, and between groups and Year levels; and the invalidity of using NAPLAN data to evaluate teachers and teaching. Missing ingredients for improved use of data are seen to be professional development in data-based decision-making, professional development in assessment literacy, and quality assurance through moderation.

Impact of NAPLAN

Improvements in learning

The overall purpose of NAPLAN is to improve student learning in literacy and numeracy, whether through ground-up action in the classroom using NAPLAN data to improve teaching in key areas or, when combined with other data, for addressing individual student needs, or through system and school level accountability driving improvement. International research literature identifies the core assumption that accountability through testing will drive improvement, although little international literature has identified learning improvement. Australian research literature is also limited on the impact of NAPLAN on teaching and learning improvement (Lingard et al., 2016), although this has not necessarily been the focus of Australian research studies. A number of Australian researchers have indicated improvement in student learning in broad curriculum aims (Brennan et al., 2016; Hardy, 2014, 2017; Harris et al., 2013; Kerkham & Comber, 2016; Thompson, 2016). Where literacy or numeracy improvement or implicit improvement has been noted in case study research, NAPLAN provided the catalyst for program development and closer collaborative examination of pedagogy in conjunction with a range of other diagnostic or assessment measures (Brennan, Zipin, & Sellar, 2016; Singh, Märtsin & Glasswell, 2015). It may also be that NAPLAN-induced improvement is occurring for specific cohorts of students, whether considered by location, according to language or cultural background, identification with Indigenous culture, disability, or different levels of achievement including students at-risk. The impact of NAPLAN for specific cohorts of students was one issue in Term of Reference 4, with these taken to include students with diverse language or cultural background, who identify as Aboriginal or Torres Strait Islander, or with disability. This is explored in a later section.

Overall, focus group participants observed that NAPLAN had led to learning improvements over time, but did not pinpoint domains of testing or Year levels where most improvement was evident in their school or

regional data. NAPLAN may not be sufficiently sensitive to demonstrate relative learning improvements. However, one impact of NAPLAN has been recognition that educational accountability is desirable in some form, and, relatedly, consensus that literacy and numeracy are essential skills for all learning.

> I don't think we should apologise for the fact that literacy and numeracy actually have to be a priority for every young Queenslander. It's on that that we then build their knowledge around HASS and science and the other learning areas. Without those basic building blocks of literacy and numeracy, it's very difficult to access any component of the Australian curriculum.

As noted in *Purpose* and *Value* of NAPLAN, many interview and focus group participants noted the impact of the early NAPLAN data, the "wake-up" call, leading to renewed focus on literacy and numeracy achievement. NAPLAN outcomes indicate statistically significant improvement for Queensland student outcomes from 2008 to 2018 (ACARA, 2017, 2018c). The evidence is that improvement as measured through NAPLAN may be plateauing from approximately 2012 (see, e.g, ACARA, 2017a, 2018), reflecting a trend reported for accountability measures by the noted education measurement theorist Linn (2000), not only in Queensland but across Australia. An ongoing concern from NAPLAN data is Writing, seen as declining across years of testing and across Year levels.

While most interviewees or focus group participants did not comment specifically on improvements in learning due to NAPLAN, apart from the comments on the impact of the original data, a small number identified overall improvement in Queensland outcomes. Some noted regional improvement against other regions, in response to the need to lift their performance. Several noted that improvement occurred in NAPLAN domains at the school level where teaching was focused, predominantly in Reading, through introduction of a school-wide instructional program in response to overall school NAPLAN outcomes, and the need to redirect teaching to other areas once those gains had been achieved. NAPLAN improvement may lie with pockets of practice, some participants identified that they used My School to identify similar schools with better NAPLAN outcomes to explore strategies that could lead to improvement.

While the initial response to NAPLAN reported by interview and focus group participants was use of NAPLAN outcomes as a "stick" to drive improvement, these participants noted the change in policy directions over the last four to five years to emphasis on schools' (i) focusing on teaching the whole Australian Curriculum, not the "NAPLAN curriculum", and (ii) focusing on level of achievement (LoA) data. Therefore, a longer-term impact of NAPLAN has been redirection of policy to learning within the Australian Curriculum and school-based assessments. The policy change reflects change in focus from the "what" to improve to the "how" to improve. The stated expectation was that NAPLAN test outcome improvement will follow.

> [in] my school we're at the point now we don't really … do NAPLAN preparation. We know that if we teach the curriculum really well and pay attention to looking at the data and making sure our program is of high quality we don't need to do NAPLAN preparation.

> We teach the Australian curriculum. This is one measure that is linked to assessment of Australian curriculum content and skills that students should have, and the test preparation should be, you know, the notion that they may do a practice test to be familiar with that.

[It] isn't about NAPLAN, it's actually about the implementation of the curriculum and if you implement that right, NAPLAN whatever is a diagnostic. … we're talking about teaching and learning, which is what we should be talking about, rather than a test.

The extent to which NAPLAN implementation has been seen to lead to learning improvement is not consistently held across Queensland school personnel. In QTU survey, 79 per cent of survey respondents identified that NAPLAN had not improved "student outcomes over the past ten years". However, 13 per cent, or one in eight, felt that it had improved outcomes, while eight per cent were not sure (QTU, 2018, p. 5).

The School Survey asked respondents to rate the effect of NAPLAN on improvements within their own school in the NAPLAN domains of testing: Reading, Writing, Spelling, Grammar and Punctuation, and Numeracy (Table 4.2). Almost half the respondents were *neutral*. Around one-third of respondents conveyed a negative (*very* or *somewhat*) view of improvement. By contrast, between one-fifth and one-quarter of respondents were positive about NAPLAN's impact on improvement in both literacy and numeracy.

Table 4.2. Impact of NAPLAN on domain performance in respondent's school (row percentages) (n = 5,814)

Domain	Very negative	Somewhat negative	Neutral	Somewhat positive	Very positive
Improvement in Reading	12.8	16.9	45.4	21.1	3.7
Improvement in Writing	16.7	19.5	41.4	19.3	3.0
Improvement in Spelling	13.7	18.4	47.4	17.9	2.7
Improvement in Grammar and punctuation	13.4	17.6	47.7	18.6	2.7
Improvement in Numeracy	12.8	17.2	47.2	19.5	3.4
Improvement in other aspects of curriculum	25.5	22.0	43.4	7.9	3.2

Views on whether NAPLAN had improved learning in the learning domains and overall differed somewhat between school leaders and teachers. While a considerable proportion of both school leaders and school staff were *neutral* about improvements in learning, school leaders were always considerably more positive, and correspondingly less negative, about NAPLAN impact on learning than teachers. For example, 40 per cent of school leaders were positive about the impact of NAPLAN on improvements in Reading, with one-third positive about impacts on other domains. For teachers, between 17 and 21 per cent were positive about these outcomes. Organisational representatives expressed rankings similar to teachers, being negative or neutral about the impact of NAPLAN on learning in all domains.

The path to NAPLAN learning improvement as a result of NAPLAN outcomes necessarily occurs through teaching. The School Survey and Student Survey both asked questions regarding improvements in teaching. Student Survey respondents from Years 7 to 10 were also unequivocal about the extent to which NAPLAN had helped them improve their learning. Seventy-nine per cent of student respondents considered that NAPLAN had not improved how their teachers taught them (*not at all, not much*), nearly 17 per cent considered it had improved *some*, while four per cent of all student respondents considered it had

improved *a lot* (Table 4.3). As for improvements in learning, Year 7 students were found to be slightly more positive than other students.

Table 4.3. How much students in Years 7 to 10 who had participated in NAPLAN felt it had helped their teachers to teach them better (n=1,341)

	Percentage
Not at all	51.8
Not much	27.2
Some	16.9
A lot	4.1
Total	100

Impact of NAPLAN and curriculum breadth; test preparation, teaching to the test

School Survey respondents rated the impact of NAPLAN on improvement in all aspects of the curriculum. Nearly half considered it has had negative impact on learning in other curriculum aspects, over half were *neutral* about impact, but only 11 per cent were positive (Table 4.2). As before, school leaders were more positive about learning impact in other curriculum aspects, one in seven felt the impact was positive to some extent. Organisational representatives were similarly negative about impact on other aspects of the curriculum.

School Survey respondents were also asked about the impact of NAPLAN on coverage of the "full" curriculum and ability to address school curriculum and program priorities (Figure 4.11). Approximately one-third were *neutral*, six per cent were positive about curriculum coverage, and thirteen per cent were positive about curriculum and program priorities. However, over 50 per cent were negative about these impacts, one-third seeing the impact as *very negative* for both. Views of school leaders and teachers differed once more, with school leaders being somewhat more positive (22%) than teachers (12%) for impact on the full curriculum and program priorities, but half or more than half of respondents in both groups reporting negative impact.

Figure 4.11. Impact of NAPLAN on full curriculum program priorities and resource allocation.

Institute for Learning Sciences and Teacher Education 2018
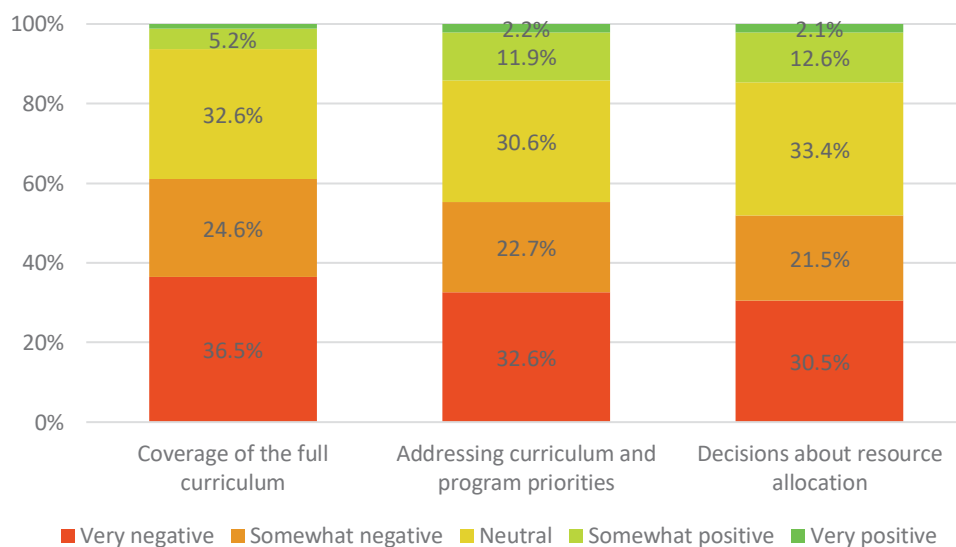
Few comments were made in interviews and focus groups regarding the impact of NAPLAN on resourcing or impact of NAPLAN on delivery of school programs and strategies. Positive comments were made in interviews and focus groups that NAPLAN data were a way of directing resources to areas of identified need.

Some participants identified the occurrence of such impact, with many schools reported to have

> either formally named or what is actually programs in place for the specific purpose of improving NAPLAN results, as opposed to programs focusing on literacy and numeracy … driven by what people perceive to be the content of NAPLAN or what's required to be taught in class in order to be successful on NAPLAN. … jobs [are] advertised in schools for … NAPLAN improvement coordinator.

Many participants commented on the narrowness of the curriculum domains assessed by NAPLAN, in terms of both constructs of literacy and numeracy, and the broader curriculum. NAPLAN was identified as testing only aspects of the curriculum, both in literacy and numeracy, and further in terms of Australian Curriculum goals and 21st century capabilities. When NAPLAN content is prioritised or indicated as valued, schools and teachers, and students will redirect attention to these areas of learning.

> … there's that incongruence at the moment between what are the pedagogies of the 21st century, what are the skills and attributes we want our learners to have when they enter the workforce and what [we are] actually testing in NAPLAN.

> … if we can get better at defining the measures that we're really after, you know, student engagement, wellbeing, critical and creative thinking, you know, the bits that we really want to know how kids operate, … a part of a suite of things that we can pull down to tell the whole story about the child.

Focus group participants therefore raised the issue of breadth and depth of NAPLAN and what has been prioritised in terms of all learning and the whole curriculum—"what I want to make sure of is the whole picture of learning".

The impact of high-stakes tests on "teaching to the test" (Brill et al., 2018) at the expense of the full curriculum is well-documented (DfE(UK), 2010; Harlen, 2005; Kramer-Dahl, 2008; Hursh, 2008; Spielman, 2017; Stobart, 2008; Stobart & Eggen, 2012), especially in the grade levels being tested (Stecher & Barron, 2001). Nearly all respondents to the QTU Survey on NAPLAN and MySchool identified that NAPLAN has become high stakes for schools, with 84 per cent identifying that NAPLAN receives a moderate to large emphasis in their schools, although the nature of the emphasis is not identified. Seventy-nine per cent of QTU respondents noted that they had felt pressured to change their teaching (whether this was a positive or negative change is not known) and 85 per cent indicated they did practice tests in schools. As one participant in this study noted

> … it's meant to be a low stakes test, but in reality, it's actually very high stakes within the community because of the use of the data, 'cause transparency is a double-edged sword. Because the data's transparent, it can then be used for purposes for which it wasn't originally intended.

The School Survey asked school leaders and teachers separately to comment on the extent to which students should be engaged in test preparation and how such preparation should occur (see *Uses of NAPLAN*). Expectations of the nature and extent of practice were somewhat similar between school leaders and teachers. Forty per cent or more of school leaders felt that test-taking strategies and practices should occur *a fair amount* or to a *substantial* amount (Figure 4.12), 55 per cent of teachers reported these practices to the same extent (Figure 4.13). More than half of respondents in both groups also reported integration of test preparation with other activities as a substantial focus of activity.

A relatively high proportion of leaders and teachers, however, considered test preparation as not important, with many rejecting the notion of focusing on elements that will be tested (which could be through embedding in the curriculum), and the overwhelming rejection of special treatment for "students at risk", perhaps interpreting this as discriminatory, rather than seeing it as targeted additional assistance for these students. Secondary teachers clearly, and not surprisingly, consider test preparation as less important than primary teachers but not overwhelmingly so. Survey responses therefore indicated more NAPLAN preparation activity than interview and focus group participants.

The emphasis placed on NAPLAN preparation in schools both in expectation and in practice was reflected in the frequency with which NAPLAN-type tests were practised. Almost half of the respondents said that NAPLAN type tests were practised at least three times during Terms 1 and 2, and nearly a quarter of respondents said that NAPLAN type tests were practised four or more times during these terms; only 10 per cent of respondents said that their schools never practised such tests during Terms 1 and 2. Practice in Terms 3 and 4 might be expected to be in the non-testing years (Figure 4.14).

Figure 4.12. School leader expectations for different types of NAPLAN preparation
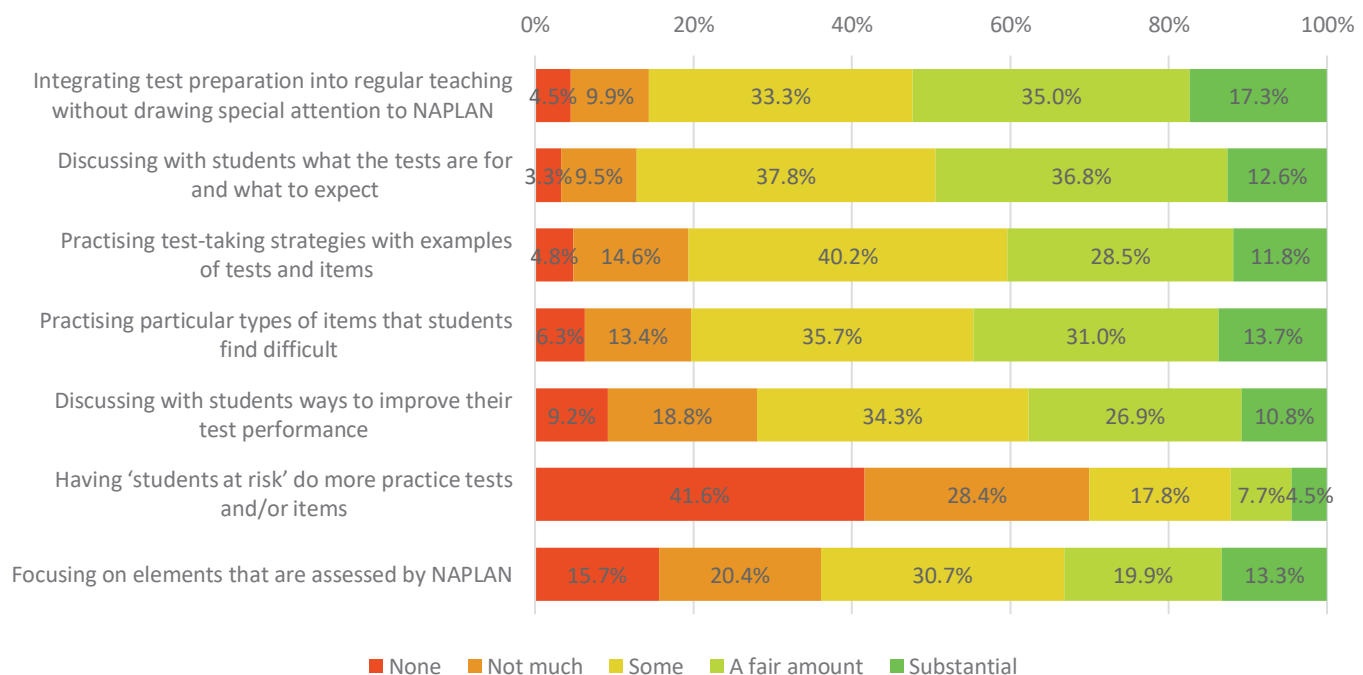
Figure 4.13. School leader expectations for different types of NAPLAN preparation
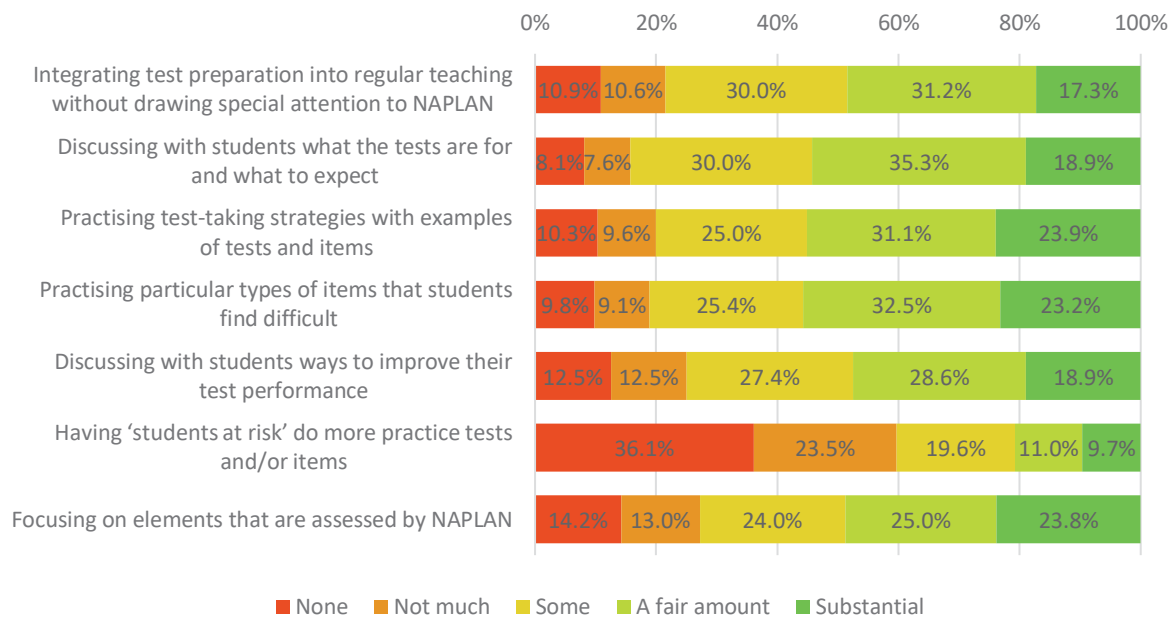
Parents in Queensland (Matters, 2018) provided a range of views on the desirability of preparing students for NAPLAN, the extent to which they are well-prepared, and time spent in schools and by teachers in text preparation. While parents were ambivalent about the need for preparation as a whole and were inclined to think students were well-prepared (*neutral* to *strongly agree*), 68 per cent felt too much time was spent on preparation by schools, and over 60 per cent considered that teachers were teaching to the test. Comments indicted that such teaching may be in the lead-up to NAPLAN testing. Time spent on test preparation and teaching to the test was noted as having impact on other teaching and learning that should be occurring. Parents noted that they should not have a role in NAPLAN preparation and few were organising private tutoring or having their child practise sample tests (Matters, 2018).

Figure 4.14. Extent of NAPLAN practice in schools
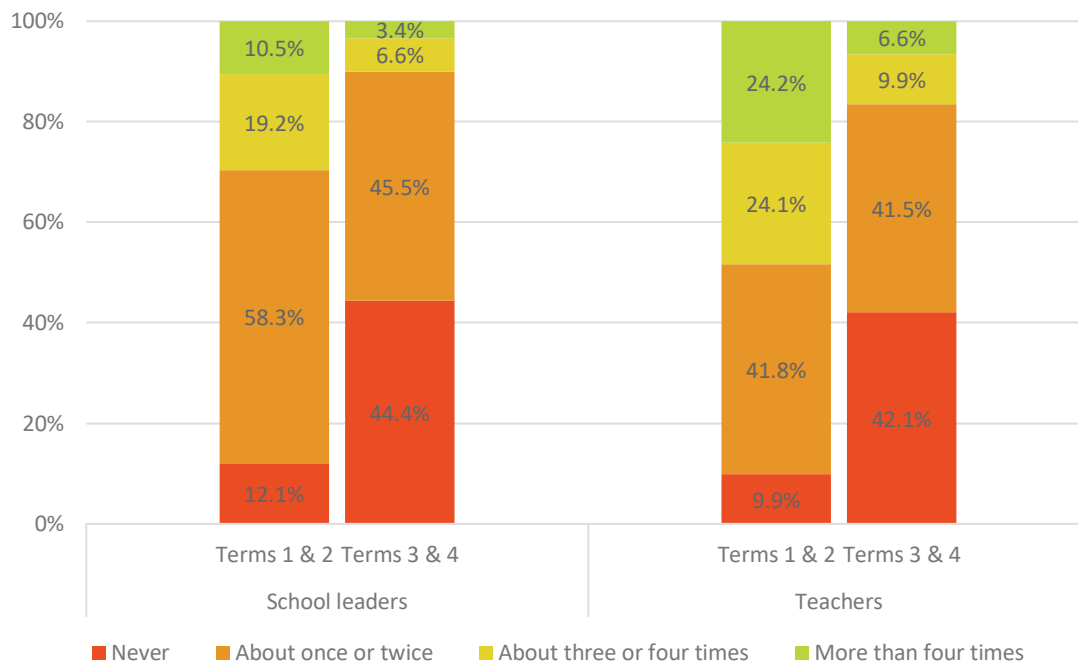
Interviewees and focus group participants noted different perceptions of the extent to which schools and teachers were engaging in extensive practice for NAPLAN including practice tests and individual items. Responses ranged from the policy directives that NAPLAN preparation and practice were to be minimal with focus on teaching the Australian Curriculum, to statements by several participants describing extreme practices in schools and resultant impact on the breadth of curriculum being taught. The general consensus was that such practices were rare, not common, with one participant saying such schools were four years behind the policy agenda.

Student Survey respondents provided information about NAPLAN preparation at school and at home. While more than a quarter indicated they did *little* or *no* practice at school, the majority reported *some* to *a lot* of school practice. Student comments also indicated that many felt too much time was spent on preparation taking away valuable class time. Many student comments stated that they would prefer to spend time learning, and that their teachers knew what they could do. Practising and preparing for the tests were specifically mentioned as using time that could be used more valuably for other activities. Student comments included comments about feeling "rushed" in the classroom.

Higher education representatives noted that one import of preparation time for NAPLAN was that schools were not willing for Initial Teacher Education (ITE) students to undertake practicums in the lead-up to NAPLAN. The sense was that schools did not want to lose class time allocated to preparation to ITE student lessons. Schools may also not want students to witness the NAPLAN preparation activities. Conversely, one higher education representative noted that some regional schools did want ITE students in classrooms during this time, as reflecting the reality of classroom experience.

Impact of NAPLAN and Writing achievement

Writing is an area that is attracting considerable focus at the national and state level. As Table 4.2 shows, almost half of the respondents were neutral for all domains, including Writing. Around one-third of respondents conveyed a negative (very or somewhat) view of improvement for all NAPLAN domains. Respondents indicated that they felt that the NAPLAN Writing test had the least impact on student improvement with 36 per cent of respondents suggesting it had a *very negative* or *somewhat negative* impact on improvement in Writing in their respective schools.

Compared to teachers, school leaders reported a more positive perspective of improvement in the domains due to NAPLAN testing. When looking at the breakdown of leaders' perspectives of both negative (very negative and somewhat negative) and positive (somewhat positive and very positive), leaders also do not feel that NAPLAN Writing is having any impact on students' Writing improvement.

However, compared to the teachers, a greater percentage of leaders (teachers: 18%, leaders: 33%) felt that NAPLAN was having a positive impact on student writing improvement. What can be drawn from both tables is that a significant proportion of teachers do not feel NAPLAN contributes to improvement or are *neutral* about its impact on student improvement.

Concern about NAPLAN writing data, or declining writing results, was a source of discussion for most focus groups. The impact of declining NAPLAN writing results was reported to be affecting teachers, and advocacy for looking beyond the data was highlighted, "[NAPLAN writing results are] demoralising for people that are working really hard and it may not necessarily be a true indication, without taking a lot of other information into consideration".

The issue of the "middle years writing is a big area that's just emerging" was highlighted, and discussions regarding the need to connect like schools where improved NAPLAN writing performance was seen as an opportunity for professional development. Engaging with like 'writing successful' schools was positioned as a professional learning opportunity for teachers to engage with successful pedagogical practices, in order to build writing expertise,

> At the moment [in] middle years, writing is a big area that's just emerging, and we will do correlations to say is it NAPLAN only or is it coming across in our other data. And if it isn't, what's happening, you know what can we learn about the schools that have got close correlations. So I think that's where NAPLAN can be very, very useful.

Participants questioned the declining NAPLAN writing results and discussed the reasons why this may be happening. One participant commented that "when we look at our writing results over the decade … it almost seems the more we test it the worse we get". Differing notions of what factors were impacting the declining standard of student writing were revealed in discussions. These included increased student "screen time" and teaching quality. In the absence of clarity about the contributing factors, some participants identified the importance of looking beyond NAPLAN data to triangulate other sources of writing data to gain a better understanding of student writing standards (discussed below) and the thorny issue of teaching quality. In the words of one participant, "then I go to, is it the quality teaching question?"

Impact of NAPLAN and student participation

Increasing rates of student non-participation in or withdrawal from NAPLAN, over recent years as well as across NAPLAN testing Year levels, are concerning at the Queensland system level. In high-stakes accountability environments, schools and systems have been shown to engage in a range of "game-playing" activities (Heilig & Darling-Hammond, 2008). One of these is to find ways to remove students from testing who are likely to have negative impact on school outcomes (Brill et al., 2018). In Australia's NAPLAN, students with disabilities and students new to Australia with only one year of English language can be exempted from NAPLAN. However, exempted students are deemed not to meet the national minimum standards and therefore have a negative impact on school NAPLAN Band profiles even though they are not included in calculations of school mean NAPLAN scores. Students who are withdrawn by parents or absent are not included as NAPLAN participants, although results for these students are imputed.

Parents have reported that schools are requesting that some low-achieving students should stay home during the test. Parents are also withdrawing students to "eliminate pressure", because of personal philosophies, or limited valuing of individual student reports (Matters, 2018, p. 19). Conversely, parents of students with special needs would prefer their children to participate in NAPLAN on the basis of inclusive practice (Matters, 2018), also noted in by focus group participants in this study. Parents do express concern about the impact of NAPLAN on students with special needs.

Interviewees and focus group participants, noting the trend in declining student participation in Queensland, also provided anecdotal evidence that (other) schools encouraged students who may not perform well to "stay away". However, more commentary was provided regarding parents exercising choice for students not to participate, from concern about their child's wellbeing, on philosophical grounds, or in response to media representations of NAPLAN (discussed below). Concern was expressed by a number of interviewees and focus group participants that often the students not participating for parental reasons were higher-performing students, often noted to be in the "upper two bands".

The culture of NAPLAN created in schools was seen as assisting in high student participation.

> … if the principal and the staff are positive about NAPLAN, if they promote it as a good thing, if they, as one of my Principals said to me yesterday, adopt the notion that everyone does NAPLAN it's just a thing we do … [student NAPLAN] participation rates seem to rise.

School principals were asked on the School Survey to provide an indication of the extent to which different reasons applied for student non-participation in NAPLAN, with multiple options provided. The major reasons noted were exemption due to disability (93%, *to some extent – very much*), parent withdrawal of a student due to student anxiety (70%, *to some extent – very much*) and parent withdrawal of a student due to disability (63%, *to some extent – very much*). Parent withdrawal due to personal philosophy was also a factor (62%, *to some extent – very much*). Student absence due to illness (53%) and exemption due to language background (37%) were less common.

Students who responded to the Student Survey were asked if they had participated in NAPLAN before they proceeded to questions regarding NAPLAN. Those who had not participated were asked if they knew why, choosing from four options (*I didn't have to do it, I was away sick, I'm not sure, My*

parents/carers didn't want me to do it) before exiting the Survey. One hundred and twenty-five (125) students had not participated in NAPLAN. Almost half of the students indicated that their parents or carers did not want them to do it, and nearly 40 per cent said they didn't have to do it. Twelve per cent were not sure, and just over two per cent were absent due to illness.

Impact of NAPLAN on teacher, student and parent wellbeing

Previous research indicates that high-stakes accountability testing can create pressure and stress on teachers, test anxiety and/or test aversion for students (see Brill et al., 2018). Thus such testing impacts on student well-being including high achieving students (see Brill et al., 2018). Findings in this project regarding teacher and student well-being are mixed. As Renshaw et al. (2013) had noted, since the introduction of NAPLAN, focus on teacher and student wellbeing has reduced. Renshaw et al. noted that few schools were recording data that related to student affective growth or wellbeing. Studies of teacher wellbeing have examined work intensification (Comber, 2012) and staff morale (Dulfer et al., 2012). Student health and wellbeing were identified as issues in the research of Dulfer et al. (2012) (also reported in Polesel et al., 2012). Negative responses to NAPLAN anxiety and behavioural issues have also been noted by Rice et al. (2016), Rogers et al. (2016), SSCEE, (2014) and Wyn et al. (2014).

Wellbeing of students was examined by Wyn, Turnbull and Grimshaw (2014) who identified the important relationship between wellbeing and learning. Their study identified that the perspectives of school personnel, parents and students were that NAPLAN contributed "significantly to anxiety and to student alienation from learning" (p. 31). A concern for many participants in this study was not the extent to which students are participating in NAPLAN, but their engagement when they were meant to be. Many comments were made that while students may be present for NAPLAN, the students themselves enacted agency in not completing the tests or randomly (or non-randomly through a pattern) selecting responses for multiple choice tests. This was most commonly associated with Year 9 students and identified as "assessment fatigue". These students were "over" NAPLAN. Many students commented on the Student Survey that NAPLAN was a waste of time and hence they did not try hard is a form of alienation from the testing.

Queensland teachers have reported that 76 per cent of students have a negative attitude to NAPLAN, although nearly 9 per cent were identified as seeing it positively (QTU, 2018). More than two-thirds of QTU survey respondents felt that NAPLAN had been harmful. Although they did not specify the nature of harm in relation to teacher and student wellbeing, their comments indicated it increased stress and anxiety for teachers, student and parents.

Matters' study (2018) of parents indicated that time spent in test preparation was a source of anxiety for students, with approximately 55 per cent identifying their children as anxious, related to fear of not doing well, although parents were also divided on the extent to which students cared about NAPLAN. Most concern was expressed by parents regarding students in Year 3, with comments that NAPLAN was less stressful as students became familiar with it. NAPLAN is seen as creating a competitive environment for schools through media publication of data and My School individual school reporting. Matters (2018) also reported parents commenting on NAPLAN use for school selection and school marketing. Several Review Phase 2 participants commented on the impact of NAPLAN for marketing and selection purposes. Directly or indirectly, such uses impact on student wellbeing.

Stress and anxiety as a result of NAPLAN may affect different groups of students differently, for example, low achieving students (Howell, 2016, 2017), students with different cultural and linguistic backgrounds (Rogers et al., 2016), students with learning difficulties (Rogers et al., 2016), and students with parents with high expectation or who have high expectations for themselves (Rogers et al., 2016). Use of NAPLAN as a school selection tool can affect students, and schools where NAPLAN is emphasised as high-stakes may have more anxiety (Howell, 2017). NAPLAN can result in low self-esteem for students (Howell, 2016; Rice et al., 2016).

Perceptions of impact of NAPLAN on teacher and student wellbeing reported by School Survey participants

Reponses to the School Survey provide evidence that NAPLAN was seen by respondents to affect staff and student well-being. More than half of respondents reported an overall negative impact on principals, a stronger effect on other school leaders (60% negative), and a much stronger effect on teachers (82% negative) (Figure 4.15).

Judgement about teacher wellbeing was related to who made the judgement. When teachers assessed their own wellbeing, 81 per cent of secondary teachers and 87 per cent of primary teachers took a negative view of their wellbeing because of NAPLAN. However, when school leaders assessed teacher wellbeing, only 67 per cent of principals took a negative view of NAPLAN-induced teacher wellbeing.

Figure 4.15. NAPLAN and wellbeing, by role in School (n = 5,814)
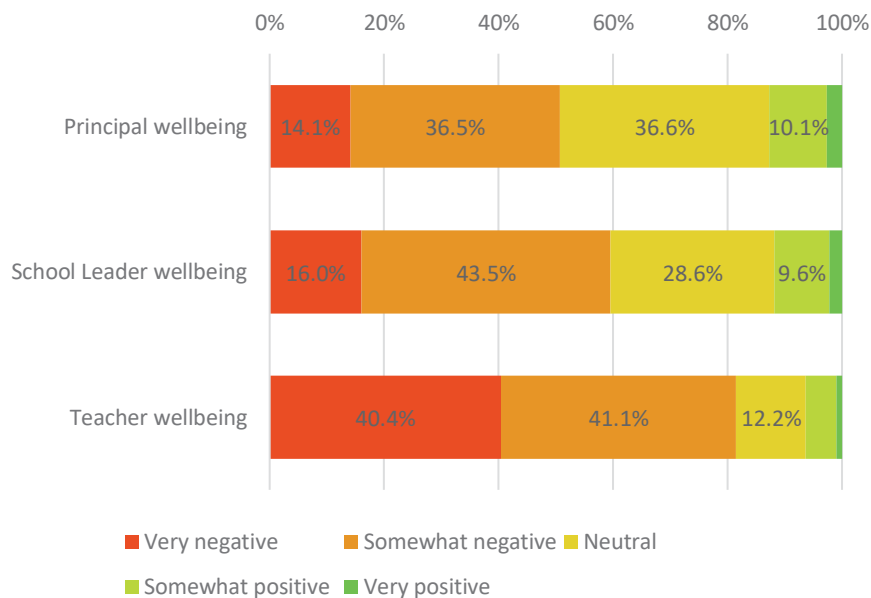
A possible reason that NAPLAN was seen as affecting wellbeing adversely is because it puts pressure on school staff. School Survey respondents rated pressure on different school stakeholders on the range *None* to *Substantial*. Sixty-nine per cent of respondents felt that pressure on principals was *a fair amount* to *substantial* while 81 per cent and 73 per cent felt similarly, respectively, for teachers and students. Judgements of the extent of pressure, as for wellbeing, depended on the role of the person making the judgement, with school leaders perceiving less pressure on classroom teachers than the teachers did.

Nearly all respondents to the Organisation Survey indicated that they considered NAPLAN to have a negative impact on school staff wellbeing, and all students, especially specific cohorts of students.

School Survey respondents also rated the impact of NAPLAN on student wellbeing, including the wellbeing of specific cohorts of students. While a proportion of respondents were *neutral* about impact, overall, respondents indicated that they considered NAPLAN to have a negative (74%, *very* or *somewhat negative*) impact on students. As Table 4.4 shows, when students in specific cohorts are considered, views on the impact on their wellbeing are much more negative, especially for students with disability (83%, *very* or *somewhat negative*) and students who have English as an additional language or dialect (EALD) (81%, *very* or *somewhat negative*). The reasons why NAPLAN is considered to have a negative impact on the wellbeing of students with disability deserve further analysis in relation to specific disabilities.

When ratings of impact are examined by school role, school leaders are less negative regarding the impact of NAPLAN on students than teachers. Only 12 per cent of principals and 18 per cent of deputy principals took a *very negative* view of NAPLAN-engendered students' welfare. By contrast, 38 per cent of secondary teachers and 38 per cent primary teachers viewed students' welfare as *very negative* and nearly 80 per cent of both teacher types viewed NAPLAN impact on students' welfare as negative (*very* or *somewhat negative*).

Table 4.4: NAPLAN and students' wellbeing

	Very negative	Somewhat negative	Neutral	Somewhat positive	Very positive
Students overall	32.2	41.7	20.4	4.8	0.9
Students with a disability	59.1	23.6	14.8	1.9	0.5
Aboriginal and Torres Strait Islanders	45.5	29.4	21.9	2.6	0.6
Students with EALD	55.0	26.2	16.1	2.3	0.5

Students in the Student Survey were not asked about their wellbeing but about their feelings about NAPLAN the last time they took the tests. They were asked to consider how they felt before doing the tests, while they were doing the tests, and after the tests. Emojis representing *very worried or sad*, *worried or sad*, *okay*, *happy*, and *very happy* were used. Thirty-six per cent of students identified that before NAPLAN tests they were *very worried or sad*, or *worried or sad*, 46 per cent were *okay*, and 18 per cent were *happy* or *very happy.* Similar feelings were reported while they were doing NAPLAN. After NAPLAN, 77 per cent of students were *okay*, *happy*, and *very happy*, indicating that students who responded to the Survey were not as concerned with their wellbeing during the experience of NAPLAN as school staff had been.

By contrast, qualitative comments provided by students provided a range of feelings about NAPLAN. While most were negative, students who may be among high achievers were more positive; low achievers are more affected. Student comments included feelings of stress and loss of self-esteem because of NAPLAN.

> Teachers have a curriculum to teach to, however the year of the NAPLAN they only seemed to teach & compare standards to NAPLAN. The NAPLAN made me so stressed that I don't believe my results

were accurate. I believe that results that are required could be obtained from the exams I already do at school.

It is extra stress and we had to stop learning the curriculum to do lots of NAPLAN practice. We also had to do several tests in preparation. I just think that especially in older grades it is unnecessary stress.

It makes me feel bad. Because I always fail.

Negative impact on wellbeing is identified as an unintended, but well-recorded, outcome of high-stakes testing environments. Such impact is in conflict with those initial goals of the national declarations for all students to be "successful" and "confident", "motivated to reach their full potential" (MCEETYA, 2008, p. 8), with a positive/resilient "sense of self-worth, self-awareness and personal identity" (p. 9). The extent to which this is realised in the context of NAPLAN as a high-stakes process is not clear.

Few interviewees and focus group participants commented on the impact of NAPLAN on wellbeing of staff or students. Some provided individual anecdotes on students who were affected but it was not raised as an overall concern. Similarly, pressures on staff were seen to relate more to overall workload than to NAPLAN specifically. The general sense was that NAPLAN had become part of the school landscape and therefore was more accepted and less anxiety-provoking.

Previous research has reported that pressure on students can arise through pressure on parents who want their children to do well, including use of "cram" schools (see Brill et al., 2018; Kwon, Lee, & Shin, 2017). Interview and focus group participants made many references to the availability of commercial products that are NAPLAN-related. A very small proportion of students (3%) indicated on the Student Survey that they practised NAPLAN, or NAPLAN-type materials a lot, at school and at home. However, two-thirds did not practise at home.

Impact of NAPLAN and negative unintended consequences identified in previous research

As noted in Chapter 2, many negative unintended consequences have been identified in previous research. These consequences emerge for the most part when test-based accountability processes become "high stakes" for systems, schools, teachers and students. Previous research (Dulfer et al., 2012; Gable & Lingard, 2015; Hardy, 2014b; Lingard, 2010; Lobasher, 2011; Polesil, 2014; QTU, 2018; Wyn et al., 2014) and Phase 2 Review data identify that although not the original intention, NAPLAN has become high stakes for key stakeholders. High stakes have been created through pressures on schools from systems to improve, in a competitive environment of jurisdictional, sector and regional comparisons, and, most notably, by media (mis)representations of NAPLAN outcomes and valued learning.

As noted, previously, whether intended or unintended, a major impact of NAPLAN has been development of focus on data literacy, at system, school and teacher level. As noted in *Uses of NAPLAN*, the focus on data literacy has played out in positive and ambiguous ways. Positively, professional development is directed to data or assessment literacy, noted in previous research as an ongoing area of need (Carey, Grainger, & Christie, 2017; Datnow & Hubbard, 2016; Pierce & Chick, 2011).

So as a system all of our staff have become much more cognisant of the importance of data driven decision making … that wouldn't have happened, there was nothing else that was going to generate that in existence prior, so that's been a really powerful trigger for that.

This may be seen as an intended outcome of the MCEETYA expectations for "world class assessments" and statements regarding "good quality" "reliable, rich data" to enable students to design programs, implement teaching strategies and improve student outcomes (MCEETYA, 2008, p. 10).

Positively or negatively, Phase 2 Review respondents identified involvement of schools in collection of a range of data on student achievements, creating some concern that the data are not necessarily clearly aligned with the Australian Curriculum or integrated within coherent theories of learning and progression.

The shift in the discourse to "A to E" levels of achievement may have led to an unintended consequence with potential impact on equity within schools. The Melbourne Declaration (MCEETYA, 2008) focuses on equity for all students in combination with goals of excellence in education. Framework principles for the Australian Curriculum identify the need for "high expectations" for each student (ACARA, 2012, p. 10). Interview and focus group participants referred to A to E outcomes generally at times, as the discussion under *Uses of NAPLAN shows,* the focus is on students achieving As and Bs. Past research has noted that one of the impacts of high-stakes test-based accountability can be focus on students at critical junctures or borderlines to raise some students into higher levels, neglecting other students (Bew, 2011; Brill et al., 2018; Jennings & Dorn, 2008). As noted, the discussion by participants regarding the focus on students in high achievement bands was not accompanied by discussion of efforts to raise the achievement of all students, or focus on students with lower achievements or students at risk.

Notwithstanding the policy directions to schools to focus on school assessments and levels of achievement (the "A to E"), there is concern that schools and teachers may be overvaluing external "hard" data forms, adopting a culture of performativity and becoming responsive to numerical representations of their work (Mockler, 2013; Ball, 2003) with the potential to weaken teacher knowledge and professional identity and assessment skills and judgement. An "avalanche" of data is not necessarily used effectively to enhance teaching and learning (Renshaw et al., 2013). Privileging measures of student performance over other areas of curriculum can have a negative effect on overall quality of teaching and learning (Ball, 2003; Brill et al., 2018), serving to increase distrust of teacher profession at the system level (Mockler, 2013; Ranson, 2003). NAPLAN can shape the nature of schooling and educational practice (Gorur, 2016; Hardy, 2015). Interview and focus group participants highlighted the importance of context in interpreting school NAPLAN data and student learning. External measures will not increase understanding of contextual factors influencing classroom learning, assessments and outcomes. Discussion with respect to early target setting noted the shift from "what" to "how", and one interviewee commented that reference to As and Bs was "shorthand". However, while references were made in interviews and focus groups with respect to teaching the Australian Curriculum and hence achieving NAPLAN outcomes, few references were made to improving overall quality of teaching in classrooms and quality of student learning.

While the discourse provided by Review participants identified a shifting focus towards school assessment, levels of achievement and triangulation, discussion under *Uses of NAPLAN* shows that how NAPLAN data

are being used to triangulate with school assessment data, whether in conjunction with or as moderators of, is not clear.

Impact of Media and My School on NAPLAN

Media "league tables" were identified by many respondents as the impetus for much of the undesirable effects of NAPLAN and suggested that the publishing of NAPLAN results needed to be stopped. The tables were described as divisive and highlighting the performance of schools in which students consistently performed well (cruising schools) rather than those schools in which students showed significant growth over time (upward trending schools). Media reporting affected school reputations, staff morale and student wellbeing. No positive comments were made about media. The media discussion of NAPLAN results was seen to focus too much attention on a small range of the curriculum; criticised as affecting teacher, student and parent morale resulting in unnecessary anxiety and stress; and going against principles of effective assessment practice used to inform and progress learning. In addition, the release of results to the media before schools had time to analyse their own results washbackas an identified issue by some respondents. The excessive time and resources spent in preparation for NAPLAN was unsurprising to some respondents given the publicity of the results and a desire to have students so familiar with the test that it was no longer perceived as a threat. Schools' concerns related to the pressure associated with the "naming and shaming" that occurred through negative media, particularly when data were misread or only part of the story of the school was told; when parents used the My School data to infer the quality of school, particularly when inappropriate comparisons with "like" schools that were considered not alike occurred; or when school reports were in the "red" without consideration of context. My School was frequently referred to as a misrepresentation of many schools. In particular, this was related to ICSEA as the basis of like-school comparison. In addition, the comparison of data over years, including the tracking of a cohort across years was described as an inaccurate measure of performance: "the school I currently work in has approx. 50 per cent change in students in a cohort when tracked from Prep to Year 6".

Although NAPLAN was recognised as an accountability measure, the overall perceptions of participants were that media commentary created the high stakes nature of NAPLAN for school staff and students. Education authorities and school staff would prefer to manage NAPLAN as a part of everyday business and a tool providing data to be used for improvement purposes.

Media representatives were seen to misuse NAPLAN data in order to sell papers while demonstrating lack of understanding of the nature of the data. Comments were made that if NAPLAN results are to be available publicly, through government agency provided lists, or My School with individual school comparisons, education authorities should take responsibility for creating the "narrative" around the data, proactively "permeating the discourse". Comments directly attributed increasing nonparticipation of students in NAPLAN to parents' perceptions formed through media coverage.

Summary

There has been limited research regarding the impact of NAPLAN on improvement in literacy and numeracy. One particular impact has been a widespread recognition of the need for some form of accountability. Another has been the recognition of literacy and numeracy as essential skills. Some research points to NAPLAN as a catalyst for program improvement in schools and for particular cohorts of students. However,

NAPLAN may not be sufficiently sensitive to demonstrate relative learning improvements. Improvement may lie in "pockets of practice" in regions and schools rather than overall. Early gains in national and state performance may be plateauing.

The longer-term trend has seen a redirection of policy and practice towards focusing on the whole curriculum and focusing on school-based assessments. There is also a change of focus from "what" to improve to "how" to improve, with NAPLAN having a lower profile within a broader context.

There is much difference of opinion among school personnel about whether NAPLAN has impacted positively or negatively on the several domains of literacy and numeracy and the curriculum. The dominant opinion was "neutral" (by almost half). As in other matters, school leaders were slightly more positive than teachers. Organisational representatives expressed similar opinions. Students overwhelmingly thought NAPLAN had no impact on their learning, and that the quality of teaching they experienced had not changed.

The general view among school personnel and organisational representatives was NAPLAN has had a negative impact on learning across the curriculum—distorting curriculum coverage and priorities by its narrow focus. Again, school leaders were slightly more positive, though still in general negative.

One expected impact of NAPLAN was attention to preparing students for the tests. A relatively high proportion of leaders and teachers considered test preparation as not important, though a majority supported it (with secondary teachers less so than primary teachers, as to be expected). Surprisingly, there was an overwhelming rejection of special attention for "students at risk" (perhaps viewed as discriminatory rather than addressing individual needs).

The amount and kind of practice testing varied substantially across schools. Some would seem excessive, but a more balanced approach seems to be emerging. Students do not like too much time to be spent on NAPLAN preparation. The period just before NAPLAN testing is the most intensive for test practice.

Increased non-participation in NAPLAN is concerning at system level. The patterns of participation, exemption, withdrawal and non-participation, and the reasons given, are complex and can distort the reported school results. Much appears to depend on the NAPLAN culture created by the school.

Wellbeing of students is a significant concern in itself and also for its impact on learning. NAPLAN can have a greater impact on some students than others (younger, low-achieving, different backgrounds, learning difficulties, disabilities, special needs). School personnel rated the impact of NAPLAN on student wellbeing as negative for students overall and more so for students with disability, Indigenous students and students with English as an Additional Language or Dialect (EALD). Students indicated little concern for their wellbeing, though this could depend on the amount and kind of pressure they felt and their "fear of failure".

A major impact of NAPLAN has been the take-up of supplementary forms of assessment, particularly standardised diagnostic tests—an anticipated privileging of test data in the absence of confidence in assessment. How and how effectively such tests are being used is unclear. There is indication that school assessments are also valued but there is some distance to travel with this.

A major impact of NAPLAN has been through the misuse of NAPLAN data by the media, which has constructed NAPLAN as high-stakes. The need for the education system to take charge of the narrative about NAPLAN and NAPLAN results was noted.

NAPLAN and Students from Specific Cohorts

Perceptions of NAPLAN and students from specific cohorts

All participants in interviews, focus groups and through the School Survey were asked through a direct question or prompt to comment on their perceptions of NAPLAN in relation to the experiences of students from specific cohorts, including students with English as an Additional Language or Dialect, students who identify as Aboriginal or Torres Strait Islander, and students with disabilities. As the previous analyses of the School Survey revealed: (a) teachers' expectations in relation to preparation were substantially lower for "at-risk" students, with 70 per cent of respondents rating their levels of expectations as *none* or *not much* and 60 per cent of teachers actualising this expectation; (b) teachers' perceptions of student wellbeing in relation to students from specific cohorts were substantially negative, with 74 per cent of respondents rating the impact of NAPLAN on student wellbeing as being *very* or *somewhat negative*; (c) school leaders' perceptions of student withdrawal in relation to student from special groups was high, with 74 per cent of respondents suggesting disability as a significant factor affecting student participation; (d) teachers and school leaders perceive NAPLAN to be of limited benefit to students with disability as well as students with EALD and Indigenous students—particularly with young children who have not yet acquired the necessary English reading skills—describing the assessments as being culturally biased; (e) teachers and school leaders perceive the time limits prescribed in the assessments as being a hindrance for all students, but specifically for children who have learning disabilities and who, with greater time, could potentially demonstrate their knowledge and skills; and (f) in relation to the preceding notion, teachers and school leaders indicated a preference towards NAPLAN catering to students' individual learning needs (i.e., in accordance with everyday classroom practice), as extra time provisions alone were ineffective for students with specific learning requirements, for example, students with dyslexia who require "the option to access the reading test aurally and the option to respond orally".

In response to the direct question or prompt, reflecting ToR 4, interview and focus group participants overall provided very few responses in relation to the NAPLAN and students from specific cohorts, although one commented that with NAPLAN, "our disadvantaged have become the most disadvantaged". Some indicated they felt too removed from practice—"I don't have enough experience in it. I don't feel comfortable [commenting on it]". Many participants noted that exemptions are available for eligible students. The existing design and outcomes, when considering the support needs of students with disability as well as students with English as an Additional Language or Dialect and Indigenous students, negatively affects students' wellbeing and self-image—"what [it] does to those kids' own self-image as learners and what it does to their engagement levels I think are of major concern". NAPLAN Online was observed as having the potential to better facilitate the experiences of students from special groups, particularly as a greater number of disability adjustment codes are added over time. In addition to these comments, participants criticised the existing focus on National Minimum Standards (NMS) as the target for Indigenous students achievement (i.e. the Closing the Gap agenda) as concerning when the predominant rhetoric

within interviews and focus groups noted both that the existing low-expectations set by the NMS (e.g. "too low") and the policy focus has shifted to "As and Bs". However, monitoring the achievement of Indigenous students NAPLAN outcomes is important in order to provide them a "voice".

> [Indigenous students' performance is] big area for use in terms of closing the gap and because we have such a high Indigenous population. So, it's an area of focus … those results often direct our targeted spending or financial delegations for future piece.

> …if you look at [Indigenous] kids in the top two bands, they're not there. So, when you look at NAPLAN it basically tells you that we're a bunch of failures. Our kids are failing. It's the system that's failing to move our kids up to that top two bands. … the NAPLAN results are not forcing them to do anything to get our kids up to there. So, what's the [purpose of NAPLAN]?

> The Indigenous in our system certainly do measure up poorly in our NAPLAN data. … But again, without NAPLAN, that's still a monitoring of your evidence and progress and success with those particular groups. So, if you didn't … it's a valuable tool, even if it is highlighting this elephant in the room.

Overall, principals and school leaders offered few comments about the experiences of students with disability or students who spoke English as an Additional Language or Dialect. However, some participants described their EALD students as performing "quite well" on NAPLAN in comparison to other reading assessments, with one participant hypothesising that it may be due to the repetitive nature of the questions in comparison with other external tests used in schools. In one of these cases, however, it was noted that the parents of these children often perceived NAPLAN as a pathway for their child's "advancement in society", even though the "children may be [left] distressed".

Extending these findings, qualitative analyses of focus group interviews revealed two main, interrelated and additional findings in relation to the perceptions of Aboriginal and Torres Islander students' experiences of NAPLAN. First, sector leaders described the purpose of NAPLAN as being a tool to "get a national picture of how Australia is doing with numeracy and literacy" and to gauge student proficiency levels within these domains; both of which are directly related to the receipt of Commonwealth funding. These same respondents suggested that all parents, irrespective of EALD or Indigenous background, aspire to the same expectations for their children. Further, NAPLAN was a "one-time testing" and teachers; report cards and alternative data sources offered greater indicators of achievement. When used as an indicator for pathways, NAPLAN was seen to limit students' opportunities, particularly in relation to school selection. For example, school-sector leaders noted that NAPLAN-associated high school selection limited Indigenous students' opportunities to attend schools that might be seen as being higher quality and offering strong extra-curricular programs (e.g. music), as illustrated in the following direct quote.

> And that's what we talked about, about families having choices because they were saying in one area they've all got to go to this school, and I said, well, they don't like to because it's not a safe and culturally safe environment or school that they're going to. These kids want to go to there, but they can't.

In framing these discussions, school-sector leaders also noted the disparity in performance between Indigenous students residing in rural and remote areas versus metropolitan areas. Second, while noting that media representations of NAPLAN suggested both positive and negative messages as an overall form of gauging student achievement, and without commenting on this aspect, further comments were made about the unique experiences Indigenous students brought to school, including their cultural and linguistic diversity, and whether NAPLAN as well as schools and teachers were able to accommodate these. For example, participants noted the stress that Indigenous students feel when they participate in NAPLAN, particularly in relation to the administration of the test within a non-familiar environment, and the test items being "culturally unsafe"—that is, inappropriate or misaligned to culturally and linguistically diverse backgrounds, knowledge and understanding. One participant suggested "if you actually did that NAPLAN testing in an Aboriginal perspective, you would have children flying off the top of the scales with some of that". Thus, language was described as being a significant issue, with both students and teachers often speaking English as an additional language, including in some cases, speaking English as a third or fourth language.

Summary

The relationship between NAPLAN and students from specific cohorts including Indigenous students, students with disability, and students with English as an Additional Language or Dialect was not identified as a prominent concern in the Review Phase 2 data. Principals, teachers and school- and sector-leaders suggested that NAPLAN posed negative implications for these students' wellbeing and provided potentially limited benefits for students identified as being "at-risk" as the associated outcomes may or may not reflect their achievement capabilities. A concern arising from the data collected in Phase 2 of the Review is that the focus on improving outcomes for students to achieve As and Bs has redirected school focus from students at risk and lower achievers. This reflects the School Survey findings that few school staff were engaging in test preparation for students at risk or focused on using NAPLAN outcomes to assist these students. It exemplifies one of the noted unintended consequences of teachers focusing on students at critical boundaries to push them into higher levels of achievement, normally associated with boundaries such as meeting minimum standards or levels such as NAPLAN boundaries. This raises fundamental issues of educational equity in conjunction with excellence for all, which are key goals of the Melbourne Declaration.

NAPLAN Online

Perceived experiences of NAPLAN Online

Three hundred and twenty (320) school leaders and teachers across 178 schools, and 118 students across Years 5, 7 and 9 responded to School Survey and Student Survey questions respectively regarding their participation in NAPLAN Online in 2018. Each of these respondents was asked to compare their recent computer-based assessment experience with their previous paper-based assessment experience and to rate their perceived experiences accordingly. Overall, School Survey responses were equally distributed across *worse*, *the same*, and *better* ratings for Online experience compared with paper-and-pencil experiences. Students were much more positive about the experience, with 55 per cent saying the experience was better, and only 15 per cent saying it was worse.

Given the potential for the NAPLAN Online experience to be affected by access to major technological and human resources, it is worth noting that majority of staff across six of the seven regions reported comparable (i.e. neutral/same) or positive (i.e. better) experiences of the computer-based assessment. For instance, 76, 80 and 86 per cent of respondents across Central Queensland, the North Coast and Far North Queensland (respectively), reported positive perceptions of their Online NAPLAN experience(s). While respondents across Metropolitan regions as well as South East Queensland and the Darling Downs South West reported similarly perceived experiences (60%, 63% and 65% respectively), School Survey findings revealed that 52 per cent of respondents across the North Coast reported a negative experience with the computer- versus paper-based assessment.

One hundred and forty-three (143) school leaders and teachers provided comments on the School Survey on their 2018 NAPAN Online experience. The majority reported positive experiences with NAPLAN online. For example, many respondents described an identifiable improvement in student engagement and apparent "work ethic", an occurrence which was particularly pronounced for students with disabilities. Descriptions included "excellent", "fantastic", "far better", "smooth", "faster" and "simple to use".

Despite these positively perceived experiences, several of these same respondents also described potential concerns relating to the NAPLAN Online implementation. Several respondents noted potential issues in relation to future resource demands. For example, several commented on the time and training required for staff both to engage successfully in the NAPLAN Online implementation, and to prepare students sufficiently for participation in the computer-based assessment. Within these comments, one staff member described feeling under-prepared, while others felt too much time had been spent on preparation, describing the implementation as being time-consuming to set-up and complicated, as well as requiring considerable administration and paperwork.

Several respondents noted technology and infrastructure concerns, including issues related to technological capacity, bandwidth, computer issues, network issues (drop-outs, glitches), as well as students losing screen displays and responses. However, these were often reported as being small-scale technical problems that were easily rectified without disruption. Aligning with concerns related to infrastructure and resources, respondents also raised the issue of prospective costs associated with the adequate provision of necessary computer hardware, computers and devices. This was particularly pronounced in relation to schools with lesser-equipped IT infrastructures, who described the two-week long window of NAPLAN Online as being disruptive logistically (e.g., when computer-based instructional classes require the same computing spaces and devices, at the same time as these were required for NAPLAN Online participation).

In relation to concerns associated with technology and infrastructure as well as time-related resources, several respondents highlighted the importance of enhancing student computer proficiency, for example, to ensure typing skills do not interfere with students' capacity to write. Associated equity considerations were also raised, particularly regarding the potential advantages and disadvantages of students with more or less-advanced computer skills, along with notions of maintaining a "normal typing writing environment" to facilitate all students written progress (e.g. enabling spell-check). Specific references relating to how it

impacted students' application to the writing task and the potential barriers to a successful online writing performance were discussed in focus groups.

The preparation for the online writing task was a source of concern for some participants. Student computer proficiency was an issue highlighted in 20 comments (14%). Comments ranged from the need for students to have more preparation to be 'test ready', that typing skills interfered with capacity to write, and that writing was too difficult for students who could not type, making NAPLAN a typing test measuring typing skills adding to the stress of NAPLAN for some students. The following direct comment illustrates several of these issues:

> The expectations to prepare students for two writing text types as per curriculum in traditional pen-and-paper to ensure pedagogy writing criteria is met in effectively 1.45 terms was then magnified by having to teach computer navigation of keyboards, disabling of predictive typing features meant a focus on using enter key to paragraph, "Shift key to capitalize etc" was a big ask.

Commentary relating to teachers implementing keyboarding skills in the classroom was also explored in the Queensland AWS (Wyatt-Smith et al., 2017). Eighty-seven percent of teachers reported that they were not prepared or minimally prepared to teach keyboarding based on their Initial Teacher Education (ITE). A further 79% of teachers reported not being prepared or minimally prepared to teach handwriting. While the challenging nature of keyboarding was discussed by participants, there is a need to focus on the teaching of keyboarding as part of pedagogical practice to ensure students are able to access online writing with speed, accuracy and fluency as part of 21st Century skills. The study found that teachers need greater support and access to Professional Development to ensure greater confidence to teach these areas.

An interesting "washback effect" was discussed by a participant; the "washback effect" was the notion that the prioritisation of getting students ready for keyboarding skills would now be a focus "into Years Two and One and Prep due to the requirement of online". The AWS study also showed the limited time provided for students to use technology in curriculum areas and to compose text online. Some schools however were in fact demonstrating this practice and had taken on board implementing keyboard skills in the early years,

> We've introduced, like demand writing down on a keyboard now in, at the end of year one to get, you know, so that they're starting to get more savvy about that but I think that you're right that you build it in and if you're teaching the curriculum then don't worry but you still have to I guess, with the online …

As mentioned earlier, consistency of online platforms for external testing was beyond the scope of the Queensland AWS research project, however, digital technologies as part of writing pedagogies repertoire beyond word processing is needed. Research from the AWS data showed that, over a fortnight, 55% of students had not used digital technology when composing texts. Greater utilisation of digital technologies beyond word processing is needed as part of 21st Century skills so that teachers and students connect with digital devices as part of exercises, and in collaboration to plan, draft, revise and edit a piece of writing. A policy consideration is whether this instruction is to focus on students' capabilities in using digital text composing functions, without the in-built functions of spell checks and grammar correction.

A further concern raised by school staff related to cognitive demand, particularly in relation to factors such as time management, information processing, reduced working-out space, and branching aspects of the computer-based assessment (e.g. increasing difficulty of questions for more able students). The notion of branching was also linked to comments relating to increases in student stress and anxiety, however, it is important to note that branching and tailoring was conversely identified as being highly beneficial for students of varying abilities by principals and school-sector leaders in interviews and focus groups. This latter notion relates to the final concern, comparability—namely, concerns that computer-based NAPLAN outcomes are not comparable with paper-based NAPLAN outcomes, with implications for the reliability and validity of NAPLAN data. Some staff commented that results were much lower than in previous years— an outcome which they believe could be attributed to the differing conditions and nature of the test itself. In contrast, others described the differing conditions and nature of the assessment as providing focused questions which challenge higher performing students while also engaging lower performing students and offering greater diagnostic information.

A further related concern was noted in relation to understanding and interpreting the Online data output in view of structural and design differences. However, as is discussed in the following section(s), similarly raised concerns in interviews and focus groups were often allayed by accompanying colleagues (i.e. other participating group members).

Interviews and focus group participants who had participated in the NAPLAN Online implementation described their 2018 experience as being positive, with many emphasising their commitment to building a culture of "excitement", "enthusiasm" and "pride" about participating in a new computer-based assessment format within their school(s). They highlighted the apparent increased level of "enjoyment" and "engagement" displayed by students during and after the testing process, as well as the "seamless" delivery of the assessment, despite concerns about "managing the process" or uninterrupted service. They made statements such as "I was very proactive about promoting it … we created a sense of enthusiasm", "my whole school embraced it with enormous positivity and a great sense of pride", and "it was absolutely seamless every day that the children sat down for the tests".

Interview and focus group participants, similar to the comments provided by School Survey respondents, offered mixed views about the differing conditions and nature of NAPLAN Online. For instance, while some participants spoke less favourably about the prescribed sequencing of the components of NAPLAN Online as well as the associated time constraints (e.g. "the sequential nature of the test was the problem for us"), others perceived this approach as offering a higher level of flexibility for schools, allowing them to "stagger starts" and spread-out the test across the two-week period. Participants predominantly offered positive comments about the "adaptive-testing" design, suggesting that "people liked the branching" as it offers students with lower-and-higher abilities an opportunity to be stretched in a way that hasn't previously been possible with paper-based assessments. Thus, while some participants recounted negative experiences associated with higher-achieving students being stretched too far (e.g. "[those] branched to the higher test were getting quite stressed because they weren't finishing"), they predominantly concurred with the perceptions of School Survey respondents:

… the online component's a little bit different in the sense that it's more tailored towards the student's journey through the testing, such as reading, and they go through different testing phases. … I think it's more authentic and tailored to the student need.

… some students extended way beyond what we felt the typical 10 out of 10 type of test that NAPLAN allows you to have … [for example] a student … who had received full marks in Year 3 in maths; in Year 5 he actually completed successfully many Band 10 questions … A couple of Year 3 students who have diagnoses including autism actually engaged better than we ever imagined with the one question on the screen at the time, the multimodal focus to be able to repeat, to hear the audio so two of them sailed the whole way through and we believe that [one] would not have … would have really been colouring just dots on the paper test before because he was told he had to colour a dot in every question so it did change so that was successful.

… our little guys with disability … can listen to the instructions repeatedly [using headphones] without everyone knowing, that's the other thing. So, there's no public humiliation. They can go back and forth between screens; [and] the computers or iPads are engaging.

Several focus group participants described challenges and concerns in relation to NBN connectivity, capacity and associated resources, including appropriate access to the internet (e.g. "in a regional centre we only just managed to get NBN"), adequate computer facilities to accommodate nine-days of testing (e.g. "that's going to take out everyone of my computer labs and … laptop areas and yet I have senior subjects that are entirely online"), and access to financial resources in order to accommodate inadequate computer facilities, particularly in schools that do not prescribe to a BYO program (e.g. "finance-wise to build in another room full of computers, there's $30–50,000 [needed] to do that"). However, in relation to connectivity issues and computer "glitches" several participants commented that despite their concerns, very few challenges were encountered. These same respondents also described enacting techniques to account for potential internet capacity issues (e.g. "staggering the starts") which served to facilitate the overall process.

Equity concerns were raised by interview and focus group participants. They noted increased potential for students in rural and remote communities to experience significant internet connectivity issues. They were also concerned for students from lower socioeconomic backgrounds and/or Indigenous students with limited access to computer devices at-home, thereby limiting their opportunities to improve computer literacy. It was however noted in the QATSIETAC focus group that Aboriginal and Torres Strait Islander students were tech savvy regarding mobile phones, with one participant suggesting, perhaps the need was for a "NAP-App" or a "N-APP-LAN".

Summary

While only a small number of schools in Queensland, and a small number of participants in Phase 2 of the Review participated in NAPLAN Online in 2018, overall, a majority of these students, principals, teachers, school-and sector-leaders have described positively perceived experiences of NAPLAN Online in 2018. While concerns regarding connectivity, resourcing, infrastructure, comparability and equity issues were raised, there is perceived potential toward improving the future NAPLAN experiences of all students.

Improvements to NAPLAN

A frequent comment from all of the groups that participated in this study was that it time for change. Comments ranged from discontinuing NAPLAN, to modifying NAPLAN, to reconstructing NAPLAN, to introducing other possibilities. In the Queensland Teacher Union report (QTU, 2018), the overwhelming majority (over 90%) of teachers were unsupportive of NAPLAN in its present form and thought that is "time to have a comprehensive review into NAPLAN and national standardised tests" (p. 14).

There was a strong view that some form of accountability is necessary to ensure high standards across the school system and provide benchmarks for improvement. The question is how to realise this without some of the misuses of data and undesirable consequences of the present system. Some suggested that greater separation of purposes is necessary for a successful system that provides useful accountability but also supports the teaching and learning of all students in all aspects of the curriculum. Others sought a redirection of resources away from national testing towards teaching and learning support, noting that the current expenditure does not appear to be cost effective in term of outcomes

Critiques and suggestions

Several critiques of the current NAPLAN testing were offered. The most prominent of these concerned: NAPLAN narrowness; online testing; the Reading test; the Writing test; and Year levels tested.

NAPLAN narrowness: The narrowness of the constructs current assessed by NAPLAN is widely seen as a problem when NAPLAN is used for accountability purposes, producing invalid judgements about system and school quality. Attention is seen as needing to be given to the full range of curriculum goals, including 21st century learning goals, as represented in the Australian Curriculum and the Melbourne Declaration.

Online testing: NAPLAN online was considered in general as a positive move, with potential for providing differentiated data that more accurately represent each student's learning in the areas assessed. However, various concerns were seen as needing attention: lack of readiness and equity for at least the immediate future; accessibility to computers; internet connection; keyboard skills—the confounding of targeted knowledge and skills with typing ability, particularly for young students; the current design of the reading tests that provided reading pieces in randomised order rather than increasing level of difficulty; the "poor quality" of the current format; and the difficulties some students experience—such those with epilepsy, who cannot spend great amounts of time in front of a screen.

The overall consensus was that any assessment of Year 3 Writing, especially, should continue to be through paper-and-pencil format, including concern about potential "washback" effect on Prep–Year 2 writing activity. There was particular concern that emphasising online writing in the early years was antithetical to good practice and cognitive research findings of the importance of handwriting as a mechanism for learning.

Reading test: There is a widespread view that the reading test is an endurance task, producing random answers towards the end ("just colouring a bubble regardless"). One suggestion was the use of fewer but well-chosen excerpts that could reveal reading and comprehension skills more efficiently.

Writing test: The writing task is seen as a major concern. The task conditions for the writing task are considered as "alien" to normal classroom experience and good pedagogy. The current writing task is seen as encouraging coaching for time-constrained, short-duration and on-demand writing, which fails to prepare students for more realistic and authentic writing activities (necessary in many school subjects and careers, and for meaningful writing). There is the further issue created by on-line writing, where typical strategies that are taught to support comprehension, such as highlighting text, drafting and editing, and using word-processing tools, are not allowed or able to be used.

> They couldn't highlight, they couldn't annotate, they couldn't do any of those sorts of things with the online version, which they saw was a particular negative in terms of doing it online.

Some considered that the Writing task was "so out of whack" with the other components of NAPLAN that writing would be best not included. In similar vein, some suggested a comprehensive strategic review of how writing is being assessed, especially in light of apparently anomalous outcomes, and it light of its inconsistency with good practice in teaching and assessing writing. It should be noted, as discussed in Chapter 2, that, in the new assessment reforms in England for school reporting and accountability, writing at Key Stage 2 will be undertaken by teachers against a broad framework within a broad range of contexts, with quality monitoring occurring through random sampling.

A further comment is that placement of the Writing task first in the set of NAPLAN tests places considerable pressure on students, especially as it is the most anxiety-provoking component of NAPLAN.

Year levels tested: Concerns were expressed about whether the current testing Years are the most appropriate. Suitability of Year 3 is challenged, in terms of age of the students, its foreign-ness to usual classroom experience and help, and the endurance required (as a whole, not just for Writing as previously noted). Suitability of Year 7 is also challenged, because of the shortness of time students have been with their (secondary) school and teachers, as well as the limbo status of Year 6, which is viewed as a follow-up Year to Year 5 testing but as "tricky territory" for teacher ownership and preparation for Year 8. One suggestion that seemed to have some support was national testing for accountability monitoring of primary and secondary schooling only in Years 4 and 8.

Sample testing and formative assessments

A key suggestion from several sources is to move from census testing to sample testing for literacy and numeracy, as is the case for other components of the National Testing Program. One enthusiast suggested that sample testing could be conducted at the end of every Year level but most did not comment on which Years would be tested and how the sampling would be done. However, the general view was that this would allow a clear separation of accountability purposes and diagnostic purposes. It would also address the issue of media portrayals of school quality through league tables. The change assumes development of more comprehensive assessments for use in schools together with appropriate support for their use. Such assessments would be used diagnostic and formative purposes within schools and classrooms; for monitoring purposes, schools could upload summative reports based on these assessments at their discretion or at appointed census date. Quality assurance processes would need to be developed.

Such an agenda would bring formative assessment (assessment for learning) to the fore, as international research recommends and as the Melbourne Declaration supports. It would also resolve a problem experienced by teacher educators, who find that the current emphasis on NAPLAN as a summative measure for accountability purposes undercuts their capacity to develop teacher graduates with broad assessment knowledge and skills that can be used for formative purposes, as required by the *Australian Professional Standards for Teachers (Standard 5)*.

Attention was also drawn by one interviewee to implications of the Growth to Achievement Report (Gonski, 2018) which promotes a formative assessment agenda. It was suggested that this implies development of formative processes that can support teachers in their daily engagement with students. It was thought also that the resulting repository of evidence might also be used for more general but less high-stakes monitoring within and across schools. The promotion of a formative assessment approach under "Gonski 2.0" (Gonski, 2018) was seen as needing much more clarification and development but as being a worthwhile direction for the future.

Summary

There is strong recognition of the need for some form of system and school accountability, and benchmarks for student achievement, but little support for NAPLAN in its present form. A comprehensive review that would clarify and separate accountability testing and formative assessment also has strong support. Ameliorating undesirable consequences of the current system is considered important, and may require separation of accountability and formative purposes in the national assessment agenda.

Issues that are thought to need attention in the current system include: the narrowness of NAPLAN and the need to attend to the complete curriculum; online testing issues (especially with respect to writing); reading test issues and demands; writing task issues, especially inconsistency with good practice in writing practice and pedagogy; and inappropriateness of the Year levels being tested.

There is support for accountability to be realised through sample testing in some form. There is also support for development of national resources to support formative assessment, a felt need in terms of professional standards for teachers and implications of the Gonski report Through growth to achievement: Report of the review to achieve educational excellence in Australian schools (Gonski, 2018).

Chapter 5: Terms of Reference and Key Findings

This report frames the conduct and findings of a study conducted for Phase 2 of the 2018 Queensland NAPLAN Review.

Within the Terms of Reference for Phase 2, a number of issues was identified for investigation (ToR4):

1. the value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level

2. how Queensland NAPLAN data is utilised, communicated and reported within schools, the broader education system and the community

3. expectations, understanding and use of NAPLAN by students, their families, school leaders and systems, and its importance to accountability and monitoring of students' outcomes

4. factors affecting NAPLAN participation

5. evidence of the impact of NAPLAN on student and staff wellbeing

6. the effect of NAPLAN on the ability of teachers to teach the full curriculum, school leaders to progress curriculum and program priorities, and schools to deliver on broader educational objectives

7. how NAPLAN affects specific student cohorts, including Aboriginal and/or Torres Strait Islander students

8. the differentiated experience of schools and students that participated in NAPLAN Online in 2018

9. the impact of NAPLAN on school and system resourcing

10. any undesirable consequences for students, teachers, school leaders, schools and the education system.

Phase 2 of the Review involved consultation with senior executives, middle managers, school leaders, teachers and students, and organisational representatives through interviews, focus groups and surveys. The issues within Term of Reference 4 were addressed in the data collection through consideration of seven key foci: purpose of NAPLAN; value of NAPLAN; use of NAPLAN; NAPLAN and special student cohorts; experiences of NAPLAN Online 2018; impact of NAPLAN; and improvements in NAPLAN. These key foci were used to frame the qualitative and quantitative data collection and data interpretation presented in Chapter 4. The findings and discussion presented from Chapter 4 are synthesised in this chapter to address each aspect of Term of Reference 4 and provision of Key Findings to address each of these.

ToR 4.1: Value of NAPLAN as a mechanism to support improvement in educational outcomes at the student, school and system level

The major acknowledgement of the value of NAPLAN to support improvement in educational outcomes at student, school and system level was the frequent reference in interviews and focus groups to the critical 2008 "wake-up" call for Queensland. Participants noted that prior to the introduction of national testing, Queensland had considered its literacy and numeracy performance to be adequate. Comparison with other jurisdictions reversed this belief. The introduction of an additional year in the early years of schooling gave Queensland students similar years of schooling experience to students in other states and territories before Year 3 NAPLAN. However, the "wake-up" call also gave impetus to a number of initiatives to address literacy and numeracy.

Overall, longitudinal NAPLAN data indicate that Queensland literacy and numeracy has statistically significantly improved since 2008. Major gains were made in the early years of testing as the students with extra schooling progressed through the system. Queensland outcomes have continued to progress, but, as for all states and territories, may have plateaued since 2012/3, a common outcome from the introduction of initiatives such as NAPLAN after a period of time.

The exception to any overall improved performance in NAPLAN has been the area of Writing. This is an area where concern is evident about the extent to which Writing instruction occurs in schools. Phase 2 participants, however, questioned the extent to which current NAPLAN Writing assessment and criteria were appropriate.

Key stakeholders who participated in Phase 2 noted that, in early years, NAPLAN performance data were used as a "stick", a negative driver intended to improve performance. In more recent times, the discourse has changed to NAPLAN being seen as a tool, or single piece of data, that can be used to "start conversations" about areas of improvement.

NAPLAN is identified as leading to two further intended or unintended consequences. Firstly, focus on NAPLAN data has led to greater awareness (ToR 4.2) of the value of data and evidence to guide resource allocation, programming, teaching and learning. The second consequence has been the growing acceptance of educational accountability. No Phase 2 participant indicated that educational accountability was not a desirable goal or professional responsibility.

Several Phase 2 participants also expressed the opinion that NAPLAN had been successful in raising Queensland literacy and numeracy performance. However, it was considered that in the 11th year of implementation, it may be time to evolve. The development of NAPLAN online, with capability to identify appropriate items for individual students, was seen as a positive move.

Key Finding 4.1.1: The introduction of NAPLAN in 2008 is seen as a "wake up" call to education in Queensland.

Key Finding 4.1.2: Longitudinal data provide evidence of statistically significant improvement in NAPLAN outcomes for Queensland since 2008.

Key Finding 4.1.3: While Queensland outcomes have contineued to progress, they may have plateaued since 2012/2013, as for all states and territories, a common outcome from the introduction of initiatives such as NAPLAN after a period of time.

Key Finding 4.1.4: NAPLAN Writing performance is a concern in Queensland, and nationally, and is an area where further exploration of teaching and assessment format is needed.

Key Finding 4.1.5: NAPLAN has led to acceptance of educational accountability as a necessary professional responsibility.

ToR 4.2: Use, communication and reporting of Queensland NAPLAN data within schools, broader education system and community

Education participants in Phase 2 indicated that the major uses of NAPLAN data within systems were to monitor and track school performance, and to direct resources, attention, and in more recent times, guidance, to support schools in literacy and numeracy achievement.

School participants identified that NAPLAN was used at school level to identify areas of curriculum that needed attention. Some participants identified use of NAPLAN to monitor individual student outcomes and track individual student progress over years of testing. Caution is noted in this report regarding the suitability of NAPLAN data, on their own, for such individual student monitoring, as is use of NAPLAN data as a measure of individual teacher competency. Participants did indicate the use of NAPLAN in conjunction with an array of other data as a way to monitor student progress and direct student learning.

Such integration of data at system and school level was seen as influenced by the introduction of NAPLAN. It has led to creation of a culture of data and evidence use to inform learning.

School leaders in Phase 2 identified a range of expectations for their own and staff engagement with NAPLAN data for a range of purposes, including accessing NAPLAN data, analysing trends in classroom performance and strengths and weaknesses, changing teaching strategies, and collaboration. Examination of trends in NAPLAN performance over time was identified as the most important expectation for teachers' practice. Engagement with external literacy and numeracy specialists in data analytics was principals' lowest expectation for the work of teachers. However, teachers indicated less engagement with the strategies identified by school leaders than their leaders preferred, and notably indicated limited engagement in collaboration with other school staff. Overall, teachers indicated limited interpretation and use of NAPLAN test data, frequently noted by participants as being received too late to be of benefit.

Effective school leadership through the establishment of a collaborative assessment culture in their schools was evidenced to varying degrees. Important differences emerge in the extent to which principals, senior managers and all teachers engage with NAPLAN data to examine school trends, strengths and needs, or principals and senior managers "scaffold" the NAPLAN data they determine to be useful for teacher access and application. Such differences also affected the extent to which teachers at different Year levels within schools were seen to "buy-in" or take ownership of NAPLAN results, that is, teachers beyond the test years of NAPLAN.

Evidence emerged of within and across school collaboration around NAPLAN data, and more importantly, sharing of strategies to address identified gaps in teaching and learning. It was noted that an externally-competitive environment could work against collaboration.

A major outcome in data use, linked to ToR 4.1 Value, has been the development of awareness of the value of data and evidence to inform programming, teaching and student learning. Frequent references were made to triangulation of data, with NAPLAN integrated with other data including classroom assessment, to identify learning needs. However, concern was expressed by participants and in this report that such data collection may be unprincipled, leading to the need for greater data literacy and assessment literacy for understanding and interpreting such information.

Communication of NAPLAN outcomes with the community and parents was not identified as a high priority by school leaders or teachers. Discussion of NAPLAN results with individual students was not seen as important by a large proportion of school leaders and teachers. School leaders were more likely to engage with parents about general matters such as the nature of NAPLAN and what it represented than the school's own performance or outcomes for an individual child.

Phase 2 participants, however, expressed strong opinions about how NAPLAN is broadcast to the community through media reports and creation of league tables and, to a lesser extent, My School data comparing schools with other schools. While criticisms were made with respect to the latter, and the extent to which My School comparisons were valid, many comments were made about the negative portrayal of individual schools, or all Queensland schools, in the media. These were considered to be ill-informed and inaccurate, intended only to market newspapers. Given such marketing is seen to occur, it is interesting that few participants indicated that parents made use of such data for school selection (ToR 4.3).

Key Finding 4.2.1: Systems and schools engage with NAPLAN data in a variety of ways and to differing extents to monitor student learning and direct teaching. Overall, leaders indicated higher expectations for staff engagement with NAPLAN data than teachers reported in their classroom practices.

Key Finding 4.2.2: Different levels of effective leadership to create collaborative school assessment cultures were evidenced. Key differences related to the extent to which all staff were engaged with senior leaders in examining NAPLAN data and their value for programming and student learning versus selection by senior leaders of the NAPLAN data they considered relevant to teachers. These differences further affected the "buy-in" of all teachers in a school to responsibility for NAPLAN outcomes.

Key Finding 4.2.3: Teachers indicated limited engagement with NAPLAN test data. This was often linked to delays in receiving NAPLAN data for effective use.

Key Finding 4.2.4: Considerable commentary was provided about extensive data collection in schools for triangulation, including NAPLAN data.

Key Finding 4.2.5: Overall, communication about NAPLAN with parents or students was not seen as important by school leaders and teachers. The nature of any discussion regarding NAPLAN was more likely to be about NAPLAN generally, and what it measured, than school or student performance.

Key Finding 4.2.6: While some concerns were expressed regarding the validity of comparisons of school performance on My School, strong concerns were voiced about the inappropriate use of NAPLAN data by media for commercial purposes.

ToR 4.3: Expectations, understanding and use of NAPLAN by students, their families, school leaders and systems, and its importance in accountability and monitoring of student outcomes

Participants in Phase 2 of the Review identified clear understanding of expectations for educational accountability in Queensland schools. They further identified that NAPLAN has been the dominant format through which such accountability has been expressed at system and school levels. Participants from both systems and schools identified that NAPLAN was one piece of data, collected at a point-in-time, to report on school outcomes and student achievement, a "starting point" for discussions. System and school level expectations, however, implicitly, are that school NAPLAN results will improve. School personnel self-identified high understanding of NAPLAN data.

The extent to which educators have attained high levels of understanding of data, that is, data literacy and assessment literacy, was reported to vary across contexts. It was identified as an area where professional development has been provided, and where need for professional development is ongoing. Continued development for teachers in their early careers was noted.

Participants overall reported little parental interest in NAPLAN data. Many commented that parents had not raised NAPLAN in conversations with staff about their children, or about the school in general. The extent to which parents understood NAPLAN was not raised. Others commented on parents seeking NAPLAN data for their child to provide for enrolment in a preferred secondary school. This was seen as concerning for the school and the child. Again, as noted under ToR 4.2, few parents were identified by participants as referring to My School to examine school performance or to select schools. The extent to which parents understood NAPLAN data was not a focus of discussion.

Few student participants in Phase 2 indicated interest in NAPLAN. A small number commented that they enjoyed NAPLAN and liked seeing their results, presumably students whose results were favourable. Another small number commented that they liked seeing what they could do or not do. Such information must be provided at school level, rather than through the individual report to students and their parents. More concerning were reports of students failing to engage in NAPLAN, discussed under ToR 4.5, essentially rejecting that NAPLAN has a role to play in their education that extends teachers' existing knowledge of their achievement through school assessments. The extent to which students understood NAPLAN data, apart from references to NAPLAN as a single test that may not reflect their achievement levels, was not a focus of discussion.

A further finding from interviews and focus groups was the continued use of NAPLAN data in performance indicators for middle management and school leaders. Despite challenging contexts, principals in some settings considered promotion unlikely unless their NAPLAN results improved. Similar expectations occurred at higher levels. Unfortunately, this continued negative use of NAPLAN for accountability may be

redirecting efforts of these staff from more constructive educational programming that may not lead directly to NAPLAN outcomes. Successful programming may be more responsive to local communities.

Key Finding 4.3.1: Phase 2 participants indicated that they had strong understanding of NAPLAN data.

Key Finding 4.3.2: Phase 2 participants reported little interest in NAPLAN from their parents, with exceptions when NAPLAN results were used for entry and selection to a secondary school of the parents' choice.

Key Finding 4.3.3: School staff and student participants indicated little student interest in NAPLAN.

Key Finding 4.3.4: Use of NAPLAN outcomes as performance indicators for middle managers and principals continues to highlight negative accountability uses of NAPLAN data in contrast to effective leadership practices.

ToR 4.4: Factors affecting NAPLAN participation

An evident decline in student participation in NAPLAN is a concern for Queensland education authorities. Information on student participation presented a range of findings. First, evidence from principals on the School Survey was that reasons for nonparticipation included exemption due to disability with a further small number of students absent due to illness. Exemption due to language background was evident to a more limited extent also. Principals noted a considerable proportion of students were withdrawn by parents either due to concern about student anxiety or to personal philosophy about the value and impact of NAPLAN. Some qualitative commentary was provided that schools were asking parents to withdraw students who were unlikely to perform well. However, conversely some principals noted that students withdrawn by parents were likely to be higher performing students, with their absence affecting school profiles. Some principals identified that the way in which NAPLAN and student participation was encouraged within schools, with a focus on a collaborative environment rather than heavy inducements, could lead to high participation.

An interesting outcome was observation that parents of students who identify as Indigenous and of students with English as an Additional Language or Dialect were more likely to ensure their child's participation, even if it caused distress to the student. They valued the NAPLAN process to identify their child's achievements. Further, parents of students with disability, while concerned for some of their children's affective state and self-esteem, also saw NAPLAN participation as a right.

Key Finding 4.4.1: There is evidence of a decline in participation due to parental concern for their child's wellbeing, and the extent to which they saw NAPLAN as a valuable process. The role of the media in portraying NAPLAN, in terms of school outcomes, teacher professionalism, and reported impact on student wellbeing, was seen as a major influence on parental values.

Key Finding 4.4.2: Parents of students from specific cohorts (EALD, Indigenous, learning needs) held more positive views about NAPLAN and their child's participation.

ToR 4.5: Evidence of the impact of NAPLAN on student and staff wellbeing

Mixed findings emerged from Phase 2 Review data about the impact of NAPLAN on student and staff wellbeing. Quantitative evidence from the School Survey showed that school personnel considered that NAPLAN had adverse impact on school leaders and teachers. Most negative impact was seen for teachers, especially primary teachers, over other school staff. The extent to which impact was seen as negative varied according to the role of the person making judgement, with principals and school leaders likely to be less negative than teachers. However, overall, opinion of Survey participants was that NAPLAN had a very negative effect on all school personnel and especially teaching staff. This was seen to be related to the pressure on educators overall.

Different perspectives were obtained through interviews and focus groups. Staff wellbeing was not frequently raised as a concern, with overall workload, including implementation of the Australian Curriculum, seen as being the major factor affecting staff wellbeing, rather than NAPLAN. Consensus appeared to be that NAPLAN, and educational accountability, are part of the education "landscape". However, participants noted once more that media representations of NAPLAN played a major part in affecting school and staff morale, affecting school reputations and implicitly or explicitly inferring the quality of teaching in Queensland schools is poor.

School personnel also identified concern about the impact of NAPLAN on the wellbeing of students, identifying strongly negative impact for all students overall, and especially for students from special cohorts (EALD, Indigenous, disability). Again, school leaders were more positive about the impact of NAPLAN on these students, but still with an overall negative perception of impact.

Very few participants in focus groups raised concern for individual student wellbeing as a major issue. In general, some comments were provided about individual students with anxiety, and more generally, the extent to which NAPLAN tests suited individual children. The majority of students identified on the Student Survey that they were *Okay* or better before, during and after NAPLAN testing. The proportion of students who were worried or sad was approximately one-third before and during NAPLAN testing, but less than a quarter after testing. This would still identify a considerable proportion of students who identify NAPLAN testing as affecting their wellbeing. Several students made comments indicating the extent to which they were anxious about NAPLAN testing. As noted, a considerable proportion of parents are identified as withdrawing their students from NAPLAN due to concerns with their emotional wellbeing and self-esteem.

Many Phase 2 participants also commented on the reduced engagement of students with NAPLAN testing even when they were officially participating. This was attributed to "assessment fatigue".

Key Finding 4.5.1: There is mixed evidence regarding the extent to which NAPLAN is affecting school personnel wellbeing. School Survey respondents indicated that NAPLAN has a major negative impact on staff wellbeing. However, interview and focus group participants indicated negative impact may reflect overall workload issues or media representations of NAPLAN and effect on school and staff reputation and morale.

Key Finding 4.5.2: The majority of students who participated in Phase 2 of the Review indicated that NAPLAN testing was not having negative impact on their wellbeing. However, there was a considerable proportion of students who reported negative feelings about NAPLAN. This may be affecting engagement with NAPLAN testing. It may also be reflected in the number of parents withdrawing their children from NAPLAN on the basis of anxiety and loss of self-esteem.

ToR 4.6: Effect of NAPLAN on the ability of teachers to teach the full curriculum, school leaders to progress curriculum and program priorities, and schools to deliver on broader educational objectives

A majority of participants identified that NAPLAN negatively affected coverage of the full curriculum and implementation of school curriculum and program priorities. Very few comments were made regarding these impacts. References were made to school contexts where NAPLAN preparation and teaching for NAPLAN dominated the school curriculum, especially at critical periods before NAPLAN. Concerns about NAPLAN and the full curriculum included commentary on the narrowness of literacy and numeracy constructs of NAPLAN both in the context of the richer Australian Curriculum and in the context of 21st century learning goals and the goals for the individual in the Melbourne Declaration. The impact of NAPLAN on implementation of the full curriculum and other school priorities was again ascribed to the "high stakes" nature of NAPLAN at system and school level. Again, media portrayals of NAPLAN were seen to contribute to this impact.

Data showed that most participants reported that their schools engaged in considerable time allocation to NAPLAN preparation, through a range of strategies. School leaders expected more preparation activities than teachers indicated were occurring in practice in their classrooms.

Interview and focus group discussions identified the extent to which the discourse with respect to NAPLAN and NAPLAN improvement had changed over the last four to five years. The emphasis was on teaching the Australian Curriculum which should lead to improved NAPLAN outcomes. However, this was accompanied by the focus on higher achieving students, as noted under ToR 4.1. Overall, the weight of evidence presented in Phase 2 indicates that in some areas there is still a time lag in appreciation of the changing policy direction to focus on teaching the Australian Curriculum, not narrowing of curriculum for NAPLAN improvement.

Key Finding 4.6.1: School personnel indicated both on the School Survey and through focus groups that attention to NAPLAN and NAPLAN outcomes did affect implementation of the full curriculum. Focus group comments identified impact in terms of reduction of focus on the full Australian Curriculum as well as broader 21st century learning goals.

Key Finding 4.6.2: NAPLAN was seen as representing narrow constructs of literacy and numeracy in terms of the Australian Curriculum constructs of literacy and numeracy, and English and Mathematics.

Key Finding 4.6.3: Interview and focus group participants indicated that the policy discourse with respect to NAPLAN is for schools and teachers to focus on teaching the Australian Curriculum and school

assessments. However, evidence with respect to the extent of practice still occurring in some schools, identified from School and Student Surveys, interviews and focus groups, indicates that this has not yet become embedded in practice in all regions and schools.

ToR 4.7: NAPLAN and special student cohorts, including Aboriginal and/or Torres Strait Islander students

As noted in Key Finding 4.4.2, parents of students from specific student cohorts placed value on NAPLAN as a part of their child's schooling experience. However, mixed evidence emerged regarding NAPLAN and students from specific cohorts including students who identify as Aboriginal or Torres Strait Islander, students who have English as an Additional Language or Dialect (EALD) and students with disability or special needs. As NAPLAN and students from special cohorts was a specific issue raised in Term of Reference 4, it was addressed through specific questions on the School Survey and for interview and focus group discussions. Overall, few participants raised issues regarding NAPLAN, these students and their achievement. For students with disability, predominant comments were that the students could be exempted, or that NAPLAN had a range of assessment adjustments in place for their assessments. Some commentary indicated that current NAPLAN formats may not suit these students or enable them to demonstrate their capabilities at optimal level. However, such comments were very few. Participants generally observed that within their schools, students with EALD did well on NAPLAN, especially after their first year of testing. Again, exemptions for those newly arrived with limited English language were noted.

Responses for Indigenous students were more mixed depending on the participants making the comment. Some participants with experience teaching in schools with a high proportion of Indigenous students indicated the suitability of NAPLAN formats for these students may not enable them to demonstrate their capabilities at optimal level or recognise their knowledge within an appropriate cultural context. Concern was noted about the extent to which the location where Indigenous students might complete NAPLAN was a "safe environment". The most compelling comment, perhaps, was concern that expectations for Indigenous students were not high. "Closing the Gap" goals have historically been focused on increasing the proportion of these students meeting the national minimum standards for NAPLAN, identified by many more broadly as too low an expectation for any student to be meaningful. When the discourse is focusing on higher achieving students in schools, similar expectations should be held for Indigenous students. The role that NAPLAN outcomes are playing in school selection was also raised as creating early barriers for schooling success and pathways for Indigenous students, given the likelihood in many settings that their early achievement will be lower than that of their non-Indigenous peers.

The general observation that emerged from engagement with interviewees and participants was that the shift in focus to higher achieving students may have led to reduced focus on students who may for a number of reasons be considered as students "at risk". While previous NAPLAN emphasis on the national minimum standards did lead to concentration on lower achieving students, there is a sense that these students are no longer an educational priority. One participant indicated that as the proportion of students below the standard was now unlikely to change, they were no longer a focus. Participants indicated that in Queensland since 2011, including Project 600, the focus has been on maximising students in the upper two bands of NAPLAN and, now, students achieving As and Bs. Focus on specific groups is inconsistent with the

equity and excellence goals of the Melbourne Declaration and National Measurement Framework, and reflects game-playing effects of accountability testing to focus on "bubble" students. Matters (2018, p. 26) also reported parents' perceptions that schools were using resources to "hotspot [students in the top two bands] … rather than students who really need help as they are underachieving". It is of interest, then, that School Survey participants did indicate that NAPLAN had very negative impact on the wellbeing of students from these specific cohorts yet indicated little attention to their preparation for NAPLAN and improvement in literacy and numeracy.

Key Finding 4.7.1: Overall, despite the specific prompts reflecting ToR4.7, few issues regarding NAPLAN and the achievement of students from specific cohorts, including students who identify as Aboriginal or Torres Strait Islander, students who have English as an Additional Language or Dialect (EALD) and students with disability or special needs, with the exceptions of wellbeing, were raised by participants.

Key Finding 4.7.2: Educational expectations for students who identify as Aboriginal or Torres Strait Islander stated in policy as national minimum standards were identified as too low; expectations should match those for other students for whom focus is on the upper two bands and As and Bs.

ToR 4.8: Experience of schools and students that participated in NAPLAN Online in 2018

Only a small proportion of schools, teachers and students who participated in Phase 2 of the Review had been involved in 2018 NAPLAN Online implementation. Nevertheless, their viewpoints present comprehensive indications of different experiences of the implementation. School staff were divided as to whether the NAPLAN Online experience was better than previous paper-based testing, however, more than half of the Student Survey participants found the experience to be better than previous paper-based testing, with only a small proportion finding that it was worse.

Comments provided by participants were informing, addressing a range of topics. Most notable were comments about student engagement and the advantage of the branching approach and technology to enable students with disability to achieve. Against this were some concerns that higher achieving students found some items that they were allocated too demanding with impact on their time. Not being able to see the end of the test, in comparison with paper-based forms, also impacted on time management.

School staff had different experiences with respect to technology. For some the Online experience was smooth, with few "glitches", while for others several connection dropouts and other technology issues hampered implementation, and also affected student engagement negatively. Some school personnel noted the additional expenses associated with having the IT infrastructure necessary for NAPLAN Online to occur. While capacity to conduct NAPLAN tests over a longer window was seen by several as advantageous, for others it impinged on overall school IT access, and, as especially noted in one school, on availability of IT resources over the duration of NAPLAN access for students who were undertaking courses of study online.

Two further comments provided related to the need for both teachers and students to have sufficient computer literacy to engage successfully with online testing. This was seen as an equity issue in contexts where students would have limited access to technology in either their homes or schools, or both.

Associated with this were concerns that any introduction of Writing online testing for Year 3 students could have negative effect on handwriting curriculum in earlier years, and theories that physical writing was an essential component of cognitive development and learning.

Further comments were made, as for NAPLAN Writing tests in general, that the online form of testing did not enable use of good drafting and editing strategies.

Overall, there was consensus that NAPLAN Online and branching were positive developments for NAPLAN in the future.

Key Finding 4.8.1: Experiences of NAPLAN Online were both positive and negative. Positive findings related to the increased engagement of most students, accessibility for students with disability, and ease of administration in many schools. Negative findings related to IT infrastructure and internet connectivity affecting not just NAPLAN Online implementation but other school administrative and educational activities for the duration of NAPLAN Online testing.

Key Finding 4.8.2: Further work appears to be necessary in order for teachers and students to develop sufficient computer literacy and keyboarding skills for successful engagement with NAPLAN Online.

Key Finding 4.8.3: Phase 2 participants expressed concern about the form of Online Writing assessment as well as potential impact on student handwriting and cognitive skill development.

ToR 4.9: Impact of NAPLAN on school and system resourcing

NAPLAN was viewed by many Phase 2 interview and focus group participants from the system level to inform effective allocation, based on NAPLAN outcomes in conjunction with other factors, of resources to schools in need. School participants also indicated that NAPLAN data, again with other data, could be used to identify areas of need for future development. Such development included prioritising curriculum areas for teaching as well as teacher professional development. Comments were made about the appointment in some schools of a NAPLAN coordinator, and provision of professional development that was intended to be literacy and numeracy more broadly, but in practice focused narrowly on NAPLAN literacy and numeracy test improvement. Approximately half of the School Survey respondents identified that NAPLAN had had a negative or very negative impact on school resource allocation, one-third were neutral, while one in seven reported positive impact. Negative impact related to over-attention in schools to NAPLAN preparation and resources related to NAPLAN testing, including the establishment of positions focused on NAPLAN improvement. One potentially negative impact of NAPLAN on resourcing at system and school level is the tendency to focus on one NAPLAN domain at the expense of others, frequently noted as implementation of Reading programs across the school.

Key Finding 4.9.1: Overall, perceptions were that NAPLAN has had positive impact in identifying areas of need at system and school level for further attention, and allocation of resources to schools and curriculum areas.

Key Finding 4.9.2: Some schools may be using financial resources to focus on NAPLAN, for example, through role creation of NAPLAN coordinators, or professional development implicitly focused on NAPLAN literacy and numeracy test score improvement, rather than quality teaching and learning more broadly.

ToR 4.10: Any undesirable consequences for students, teachers, school leaders, schools and the education system

As noted under ToR 4.1, the impact of NAPLAN has created a number of unintended negative consequences similar to those that have been previously identified in research literature. These have included: narrowing of the curriculum and excessive test preparation to focus on improvement in NAPLAN outcomes, rather than broader teaching and learning goals; some evidence of game-playing including requesting parents to withdraw students who may be low achieving; focus on students just below or at key junctures to improve their achievement at the expense of other students' education; use of NAPLAN as a performance indicator for senior management, school principals and teachers; and use of NAPLAN for school marketing and selection without regard for context of outcomes. Comments indicate that while many parents are nonchalant about NAPLAN and their child, others are engaging in practice and even "cram" schools, in anticipation of selection and entry to desired secondary schools.

These negative consequences relate directly to the extent to which NAPLAN is seen as an external accountability measure of system, school and classroom education quality. They indicate focus on achieving higher NAPLAN outcomes within a competitive, not collaborative, school environment. NAPLAN was originally introduced from policies to identify students needing literacy and numeracy improvement, within a framework of national monitoring of jurisdictional quality. As Phase 2 participants noted, NAPLAN became the object of learning in itself, with initial implementation creating NAPLAN as the driver of school performance and system, school and teaching and learning. Focus was on the "what" (numbers) to improve. Over the last four or five years, engagement with NAPLAN at the policy level is perceived to have changed to be more constructive, focusing on the "how" to improve within the context of the Australian Curriculum and school assessments. However, consistently throughout the evidence obtained in this study is the reported impact of the media. It is the way media portray NAPLAN outcomes, school excellence and school quality that creates the competitive role of NAPLAN across school communities. Availability of individual school data both through My School, but more importantly, through provision of data files used to create newspaper league tables, may enhance the reputation of some schools but is more likely affect many negatively. Many participants noted the need for the education sector to control the narrative used in the community with respect to NAPLAN.

A second consequence, that may be negative in practice, is evidence of overreliance on collection of a gamut of external measures of student achievement that may or may not align with the Australian Curriculum or with coherent principles of learning theory. How schools are using these to triangulate data with NAPLAN to inform teaching and learning is beyond the scope of data collected in this study. However, the evidence is that students are not only being tested through NAPLAN but through a range of measurement tools, sometimes with pre-test and post-test for three-weekly cycles. This not only impacts on teaching and learning time, rather than testing time, but also may be one factor in evidence of developing student test "fatigue".

Key Finding 4.10.1: The high stakes accountability of NAPLAN has led to a range of unintended negative consequences and practices for schools, teachers and students in schools. These include allocation of time to NAPLAN test preparation and practice, narrowing of curriculum to focus on NAPLAN elements, and focus on "bubble" students at specific performance levels, in this case, reported to be "upper two bands" or "As and Bs".

Key Finding 4.10.2: Media representations of NAPLAN create a competitive high stakes accountability environment that leads to negative NAPLAN practices.

Key Finding 4.10.3: The extent to which NAPLAN has led to high levels of test-taking in schools using a range of sources may have negative impact on the quality and breadth of teaching and learning over longer cycles.

Overall conclusion

NAPLAN implementation in 2008 created awareness in Queensland of the need to direct attention to student learning in literacy and numeracy. Over time, it has led to improved Queensland performance, in conjunction with increased schooling for children in early learning years. It has served as both a negative and positive driver of education. Current policy emphases in Queensland for schools are strong foci on teaching the Australian Curriculum and school assessment against the curriculum, with NAPLAN seen as one piece of data to inform systems and schools, and parents and the community, about student learning. However, emphasis on NAPLAN as an accountability measure at system and school levels continues to create a negative competitive environment for systems and schools, perpetuating negative educational practices in some schools. Media publication of league tables is seen as creating this environment, distracting schools and teachers from quality teaching and learning practices to suit the needs of learners in the 21st century, recognised in the Australian Curriculum and Melbourne Declaration.

Participants in Phase 2 of the 2018 Queensland NAPLAN Review were relatively comfortable with educational accountability for transparency of educational outcomes and monitoring the health of an education system. They were less confident that NAPLAN in 2018 is still achieving this goal. Ways for improvement of a 21st century-focused accountability system were noted, including the shift from NAPLAN as a census test to a sample test, similar to other National Assessment Program tests. This would necessarily reduce the creation of league tables by media and resultant impacts on practice. System and school personnel noted the need for some indicators of school performance for each school to remain accountable, but considered other mechanisms may be more suitable. Phase 2 participants identified the need to value and hence gauge educational success in all desired educational outcomes for students. They also appreciated the provision of timely data that assisted in identifying areas of curriculum and individual student learning that needed to be addressed while allowing celebrations of success. Many participants indicated that NAPLAN had served its purpose but it was time for accountability assessment in Australia to evolve.

REFERENCES

Amrein-Beardsley, A., Berliner, D., & Rideau, S. (2010). Breaking professional law: Degrees of cheating on high stakes tests. *Education Policy Analysis Archives*, *18*(14), 2–33.

Anderson, J. (2009). Using NAPLAN items to develop students' thinking skills and build confidence. *Australian Mathematics Teacher*, *65*(4) 17–24.

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9*(3), 292–327.

Australian Council for Educational Research [ACER]. (2012). *The National School Improvement Tool*. Retrieved from www.acer.edu.au/nsit.

Australian Curriculum and Assessment Reporting Authority [ACARA]. (nd). *NAPLAN has an important role in the Australian education space.* Retrieved from http://docs.acara.edu.au/resources/20150424_Reports_supporting_NAPLAN_value.pdf

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2008). *National Assessment Program–Literacy and Numeracy. Achievement in reading, writing, language conventions and numeracy: National report.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2012). *The shape of the Australian Curriculum: Version 4.0.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2013). *National Assessment Program–Literacy and Numeracy. Achievement in reading, persuasive writing, language conventions and numeracy: National report for 2013.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2014). *National Assessment Program–Literacy and Numeracy. Achievement in reading, persuasive writing, language conventions and numeracy: National report for 2014.* Sydney, Australia: ACARA.

Australian Curriculum and Assessment Reporting Authority [ACARA]. (2015a). *Guide to understanding ICSEA (Index of Community Socio-Educational Advantage values).* Retrieved from http://docs.acara.edu.au/resources/Guide_to_understanding_icsea_values.pdf

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2015b). *National Assessment Program–Literacy and Numeracy. Achievement in reading, persuasive writing, language conventions and numeracy: National report for 2015.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2016). *National Assessment Program–Literacy and Numeracy. Achievement in reading, writing, language conventions and numeracy: National report for 2016.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2017a). *NAPLAN Achievement in reading, writing, language conventions and writing: National report for 2017*. Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2017b). *National Assessment Program–Literacy and Numeracy. Achievement in reading, writing, language conventions and numeracy: National report for 2017.* Sydney, Australia: ACARA.

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018a). *NAPLAN.* Retrieved from https://www.nap.edu.au/naplan

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018b). *NAPLAN aligned with the Australian Curriculum.* Retrieved from https://www.nap.edu.au/naplan/australian-curriculum

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018c). *NAPLAN infographic.* Retrieved from https://www.nap.edu.au/_resources/Acara_NAPLAN_Infographic(V4-2).pdf

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018d). *Trend results: Mean scores.* Retrieved from https://reports.acara.edu.au/NAP/TimeSeries

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018e). *NAPLAN on paper: Information for parents and carers*. Retrieved from http://nap.edu.au/docs/default-source/default-document-library/naplan-2018-information-brochure-for-parents-and-carers.pdf?sfvrsn=2

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2018f). Preliminary [NAPLAN] results. Retrieved from http://reports.acara.edu.au/NAP/NaplanResults

Australian Education Union [AEU] (2017, August 22). *Australian Education Union opposes NAPLAN online as teachers and principals raise concerns* [Press Release]. Retrieved from http://www.aeufederal.org.au/news-media/media-releases/2017/august/220817

Australian Institute for Teaching and School Leadership [AITSL]. (2011). *Australian Professional Standards for Teachers.* Retrieved from https://www.aitsl.edu.au/docs/default-source/general/australian-professional-standands-for-teachers-20171006.pdf?sfvrsn=399ae83c_12

Australian Institute for Teaching and School Leadership [AITSL]. (2015). *Australian Professional Standards for Principals and the Leadership Profiles.* Retrieved from https://www.aitsl.edu.au/docs/default-source/default-document-library/australian-professional-standard-for-principals-and-the-leadership-profiles652c8891b1e86477b58fff00006709da.pdf?sfvrsn=11c4ec3c_0

Australian Primary Principals Association [APPA]. (2013). *Primary principals: Perspectives on NAPLAN testing and assessment*. Sydney, Australia: Canvass Strategic Opinion Research.

Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review: Assessment and learning*. Oslo, Norway: Norwegian Knowledge Centre for Education.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, *18*(2), 215–228.

Bew, P. (2011). *Review of Key stage 2 testing, assessment and accountability: Progress report*. London, England: Department for Education. Retrieved from https://www. gov. uk/government/uploads/system/uploads/attachment_data/file/18040, 1.

Biesta, G. J. (2004). Education, accountability, and the ethical demand: Can the democratic potential of accountability be regained? *Educational theory, 54*(3), 233–250.

Bishop, K., & Bishop, K. (2017). Improving student learning outcomes: Using data walls and case management conversations. *Literacy Learning: The Middle Years, 25*(1), i.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5(*1), 7–74.

Bloxham, R., Ehrich, L., & Iyer, R.fi (2015). Leading or managing? Assistant Regional Directors, School Performance, in Queensland. *Journal of Educational Administration*, 53(2), 354–373.

Boudett, K. P., City, E. A., & Murnane, R. J. (2006). The 'data wise' improvement process. *Harvard Education Letter*, *11*(4), 1–3.

Bousfield, K., & Ragusa, A. T. (2014). A sociological analysis of Australia's NAPLAN and My School Senate Inquiry submissions: The adultification of childhood? *Critical Studies in Education*, *55*(2), 170–185.

Brennan, M., Zipin, L., & Sellar, S. (2016). Negotiating with the neighbours: Balancing different accountabilities across a cluster of regional schools. In B. Lingard, G. Thompson, & S. Sellar (Eds.) *National testing in Schools: An Australian assessment* (pp. 199–211). New York, NY: Routledge.

Brill, F., Grayson, H., Kuhn, L., & O'Donnell, S. (2018). *What impact does accountability have on curriculum, standards and engagement in education? A literature review.* Slough, England: NFER.

Carey, M., Grainger, P., & Christie, M. (2018). Preparing preservice teachers to be data literate: A Queensland case study. *Asia-Pacific Journal of Teacher Education*, *46*(3), 267–278.

Chudowsky, N., & Chudowsky, V. (2010). *State test score trends through 2008-09, Part 1: Rising scores on state tests and NAEP.* Washington, DC: Center on Education Policy.

Comber, B. (2012). Mandatory literacy assessment and the reorganisation of teachers' work: Federal policy, local effects. *Critical Studies in Education*, 53(2), 119–136.

Comber, B., & Cormack, P. (2011). Education policy mediation: Principals' work with mandatory literacy assessment. *English in Australia*, *46*(2), 77–86.

Commonwealth of Australia, Department of the Prime Minister and Cabinet [CADPMC]. (2018). *Closing the Gap Prime Minister's report 2018*. Canberra, Australia: CADPMC. Retrieved from https://closingthegap.pmc.gov.au/sites/default/files/ctg-report-2018.pdf

Condron, D. (2011). Egalitarianism and educational excellence: Compatible goals for affluent societies? *Educational Researcher*, *40*(2), 47–55.

Cormack, P., & Comber, P. (2013). High-stakes literacy tests and local effects in a rural school. *Australian Journal of Language and Literacy*, *36*(2), 78–89.

Creagh, S. (2016). Understanding the politics of categories in reporting national test results. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 110–125). London, England: Routledge.

Cromey, A., & Hanson, M. (2000). *An exploratory analysis of school-based student assessment systems.* Retrieved from https://files.eric.ed.gov/fulltext/ED452221.pdf

Cumming, J. J. (2012). *Valuing students with impairments: International comparisons of practice in educational accountability*. Dordrecht, The Netherlands: Springer.

Cumming, J. J., & Dickson, E. (2013). Educational accountability tests, social and legal inclusion approaches to discrimination for students with disability: A national case study from Australia. *Assessment in Education: Principles, Policy & Practice*, *20*(2), 221–239.

Cumming, J. J., Kimber, K., & Wyatt-Smith, C. M. (2012). Enacting policy, curriculum and teacher conceptualisations of multimodal literacy and English in assessment and accountability. *English in Australia, 47*, 9–18.

Cumming, J. J., Kimber, K., & Wyatt-Smith, C. M. (2011). Historic Australian conceptualisations of English, literacy and multimodality in policy and curriculum and conflicts with educational accountability. *English in Australia, 46*, 42–55.

Cumming, J. J., Maxwell, G. S., Colbert, P., & Jackson, C. (2018). *School Survey.* Brisbane, Australia: Institute for Learning Sciences and Teacher Education, Australian Catholic University.

Cumming, J. J., Maxwell, G. S., & Wyatt-Smith, C. M. (2016). School leadership in assessment in an environment of external accountability: Developing an assessment for learning culture. In G. Johnson, & N. Dempster (Eds.), *Leadership for learning and effective change* (pp. 221–237). Dordrecht, The Netherlands: Springer.

Cumming, J., Wyatt-Smith, C., & Colbert, P. (2016). Students 'at risk' and the National Assessment Program Literacy and Numeracy (NAPLAN): The 'collateral damage'. R. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing and its effects: Evidence from* Australia (pp. 126–138). Oxon, England: Routledge.

Cumming, J., Wyatt-Smith, C., Elkins, J., & Neville, M. (2006). *Teacher judgment: Building an evidentiary base for quality literacy and numeracy education. Final report.* Brisbane, Australia: Centre for Learning Research, Griffith University. Retrieved from https://www.qcaa.qld.edu.au/downloads/publications/research_qsa_teacher_judgment.pdf

Curriculum Corporation [CC]. (2000). *Literacy benchmarks Years 3, 5 & 7.* Sydney, Australia: Curriculum Corporation.

Curriculum Corporation [CC]. (2005a). *Statements of learning for English*. Carlton South, Australia: Curriculum Corporation. Retrieved from http://www.curriculum.edu.au/verve/_resources/SOL_English_Copyright_update2008_file.pdf

Curriculum Corporation [CC]. (2005b). *Statements of learning for mathematics*. Carlton South, Australia: Curriculum Corporation. Retrieved from http://www.curriculum.edu.au/verve/_resources/SOL_Mathematics_2006.pdf

Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record*, *105*(1), 23.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). *Accountability for college and career readiness: Developing a new paradigm.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing schools use data to improve instruction for elementary students.* Los Angeles, LA: Center on Educational Governance, Rossier School of Education, University of Southern California.

Davies, J. (2012). Facework on Facebook as a new literacy practice. *Computers & Education*, *59*(1), 19–29.

Dempsey, I., & Conway, R. (2005). Educational accountability and students with a disability in Australia. *Australian Journal of Education*, *49*(2), 152–168.

Department for Education [DfE(UK)]. (2010, November 5). *Lord Bew appointed to chair external review of testing* [Press Release]. Retrieved from https://www.gov.uk/government/news/lord-bew-appointed-to-chair-external-review-of-testing

Department for Education [DfE(UK)]. (2018a). *Guidance. School performance tables: How we report the data.* Retrieved from https://www.gov.uk/government/publications/school-performance-tables-how-we-report-the-data/school-performance-tables-how-we-report-the-data

Department for Education [DfE(UK)]. (2018b). *Key stage 1 teacher assessment data collection guide*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/710196/Key_stage_1_teacher_assessment_data_collection_guide.pdf

Department of Employment, Education, Training and Youth Affairs [DEETYA]. (1998). *Literacy for all: The challenge for Australian schools*. Canberra, Australia: DEETYA.

Department of Education Queensland Government [DoE]. (2018). *NAPLAN 2018 review*. Retrieved from https://qed.qld.gov.au/programs-initiatives/education/naplan-2018-review

Department of Employment, Education, Training and Youth Affairs [DEETYA]. (1998). *Literacy for all: The challenge for Australian schools*. Canberra, Australia: DEETYA.

Dulfer, N., Polesel, J., & Rice, S. (2012). *The experience of education: Impacts of high stakes testing on school students and their families: An educator's perspective*. Sydney, Australia: Whitlam Institute, University of Western Sydney.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Earl, L., & Katz, S. (2002). Leading schools in a data rich world. In K. Leithwood, P. Hallinger, G. Furman, K. Riley, J. Macbeath, P. Gronn, & B. Mulford (Eds.), *Second international handbook of educational leadership and administration* (vol. 8, pp. 1003–1024). Dordrecht, The Netherlands: Kluwer.

Education Council, Australian Curriculum, Assessment and Reporting Authority. (2015). *Measurement framework for schooling in Australia 2015.* Sydney, NSW: ACARA.

Elliot, S., Davies, M., & Cumming, J. J. (2016). Documenting support needs and adjustment gaps for students with disabilities: Teacher practices in Australian classrooms and on national tests. *International Journal of Inclusive Education, 20*, 1252–1269. doi: 10.1080/13603116.2016.1159256

Gable, A., & Lingard, B. (2016). NAPLAN data: A new policy assemblage and mode of governance in Australian schooling. *Policy Studies*, *37*(6), 568–582.

Gonski, D. (Chair). (2018). *Through growth to achievement: Report of the review to achieve educational excellence in Australian schools (2.0)*. Canberra, Australia: Australian Government.

Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*, *15*, 598–616.

Grasby, K. L., Byrne, B., & Olson, R. K. (2015). Validity of large-scale reading tests: A phenotypic and behaviour-genetic analysis. *Australian Journal of Education*, *59*(1), 5–21.

Hardy, I. (2013). Testing that counts: Contesting national literacy assessment policy in complex schooling settings. *Australian Journal of Language and Literacy, 36*(2), 67–77.

Hardy, I. (2014a). A logic of appropriation: Enacting national testing (NAPLAN) in Australia. *Journal of Education Policy*, *29*(1), 1–18.

Hardy, I. (2014b). A logic of enumeration: The nature and effects of national literacy and numeracy testing in Australia. *Journal of Education Policy*, *30*(3), 335–362.

Hardy, I. (2015a). Data, numbers and accountability: The complexity, nature and effects of data use in schools. *British Journal of Educational Studies*, *63*(4), 467–486.

Hardy, I. (2015b). 'I'm just a numbers person': the complexity, nature and effects of the quantification of education. *International Studies in Sociology of Education*, *25*(1), 20–37.

Hardy, I. (2016). Contesting and capitalising on NAPLAN. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 98–109). London, England: Routledge.

Hardy, I. (2017). Measuring, monitoring, and managing for productive learning? Australian insights into the enumeration of education. *Revista de la Asociación de Sociología de la Educación (RASE)*, *10*(2), 192–208.

Hardy, I. (2018). Governing teacher learning: Understanding teachers' compliance with and critique of standardization. *Journal of Education Policy*, *33*(1), 1–22.

Hardy, I., & Boyle, C. (2011). My School? Critiquing the abstraction and quantification of education. *Asia-Pacific Journal of Teacher Education*, 39(3), 211–222.

Hardy, I., & Lewis, S. (2017a). The 'doublethink' of data: Educational performativity and the field of schooling practices. *British Journal of Sociology of Education*, *38*(5), 671–685.

Hardy, I., & Lewis, S. (2017b). Visibility, invisibility, and visualisation: The danger of school performance data. *Pedagogy, Culture & Society*, *26*(2), 233–248.

Harris, P., Chinnappan, M., Castleton, G., Carter, J., De Courcy, M., & Barnett, J. (2013). Impact and consequence of Australia's National Assessment Program-Literacy and Numeracy (NAPLAN)—Using research evidence to inform improvement. *TESOL in Context*, *23*(1/2), 30.

Hatch, J. A., & Grieshaber, S. (2002). Child observation and accountability in early childhood education: Perspectives from Australia and the United States. *Early Childhood Education Journal*, *29*(4), 227–231.

Hargreaves, A., & Shirley, D. (2009). *The fourth way: The inspiring future for educational change*. Thousand Oaks, CA: Corwin Press.

Hargreaves, A., & Shirley, D. (2012). *The global fourth way: The quests for educational excellence*. Thousand Oaks, CA: Corwin.

Harlen, W. (2005). Teachers' summative practices and assessment for learning—Tensions and synergies. *The Curriculum Journal, 16*, 207–223.

Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice, 16*(3), 365–381. Doi:10.1080/09695940903319745

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75–110.

Heritage, M. (2014). The place of assessment to improve learning in a context of high accountability. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 337–354). Dordrecht, The Netherlands: Springer.

Hipwell, P., & Klenowski, V. (2011). A case for addressing the literacy demands of student assessment. *Australian Journal of Language and Literacy*, *34*(2), 127.

Holmes-Smith, P. (2005). *Assessment for Learning: Using statewide literacy and numeracy tests as diagnostic tools*. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1009&context=research_conference_2005

Hout, M., & Elliot. S. W. (2011). *Incentives and test-based accountability in education.* Washington, D.C.: National Academies Press. https://doi.org/10.17226/12521

Howell, A. (2017). Because then you could never ever get a job!: Children's construction of NAPLAN as high stakes. *Journal of Education Policy*, *32*(5), 564–587.

Howell, A. (2016). Exploring children's lived experiences of NAPLAN. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 164–180). London, England: Routledge.

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline of educational equality. *British Educational Research Journal*, *31*, 605–22.

Hursh, D. (2008). *High-stakes testing and the decline of teaching and learning.* New York, NY: Rowman & Littlefield.

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. *Yearbook of the National Society for the Study of Education*, *106*(1), 105–131.

Independent Education Union of Australia [IEUA]. (2013). *Inquiry into the effectiveness of the National Assessment Program—Literacy and Numeracy (NAPLAN).* Retrieved from https://www.aph.gov.au/DocumentStore.ashx?id=7ce5ae54-f645-4ae7-901c-b572a70694ef&subId=11587

Isensee, L., & Butrymowicz, S. (2012, July 1). 'Florida teacher evaluations tied to student test scores.' *The Huffington Post*. Retrieved from https://www.huffingtonpost.com.au/entry/florida-teacher-evaluatio_n_1079758

Jennings, J., & Dorn, S. (2008). The proficiency trap: New York City's achievement gap revisited. *Teachers College Record*.

Joint Committee of American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

Joseph, B. (2018). *Why we need NAPLAN: Research report*. Canberra, Australia: National Library of Australia.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, (3), 527.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73.

Kane, M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2) 198–211.

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). *The effects of standardized testing*. Boston, MA: Kluwer-Nijhoff Publishing.

Kemp, D. (1999, May). Outcomes reporting and accountable schooling. Keynote presentation at the Curriculum Corporation National Conference. Retrieved from http://www.curriculum.edu.au/mceetya/nationalgoals/kemp.htm. Accessed 24 August 2011, no longer available.

Kerkham, L., & Comber, B. (2016). Literacy leadership and accountability practices: Holding onto ethics in ways that count. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 86–97). London, England: Routledge.

Kerkham, L., & Nixon, H. (2014). Literacy assessment that counts: Mediating, interpreting and contesting translocal policy in a primary school. *Ethnography and Education*, *9*(3), 343–358.

Klenowski, V. (2014). Towards fairer assessment. *Australian Educational Researcher*, *41*(4), 445–470. https://doi.org/10.1007/s13384-013-0132-x

Klenowski, V. (2016). Questioning the validity of the multiple uses of NAPLAN data. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 44–56). London, England: Routledge.

Klenowski, V., & Gertz, T. (2009, August). Culture-fair assessment: Addressing equity issues in the context of primary mathematics teaching and learning. In *Assessment and student learning: Collecting, interpreting and using data to inform teaching research conference*.

Klenowski, V., & Wyatt-Smith, C. M. (2012). The impact of high stakes testing: The Australian story. *Assessment in Education, Principles, Policy & Practice, 19*, 65–79. doi:10.1080/0969594x.2011.592972

Klenowski, V., & Wyatt-Smith, C. (2016). The impact of high stakes testing: The Australian story. In T. J. H .M. Eggen & G. Stobart (Eds.), *High-stakes testing in education: Value, fairness and consequences* (pp. 65–79). London, England: Routledge.

Knapp, M. S., Copland, M., & Swinnerton, J.A. (2007). Understanding the promise and dynamics of data-informed leadership. In P.A. Moss (Ed.), *Evidence and decision making.* (The 106th Yearbook of the National Society for the Study of Education, Part I, pp. 74–104). Malden, MA: Blackwell.

Koch, M. J., & DeLuca C. (2012). Rethinking validation in complex high-stakes assessment contexts. *Assessment in Education: Principles, Policy & Practice, 19*, 99–116. doi:10.1080/0969594X.2011.604023

Kohn, A. (2001). Fighting the tests. *Phi Delta Kappan*, *82*(5), 349–57.

Kramer-Dahl, A. (2008). Still an examination culture—for most: Singapore literacy education in transition. *Curriculum Perspectives*, *28*(3), 82–89.

Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice, 24*, 60–77. DOI:10.1080/0969594X.2015.1074540

Lewis, S., & Hardy, I. (2014). Funding, reputation and targets: The discursive logics of high-stakes testing. *Cambridge Journal of Education*, *45*(2), 245–264.

Lewis, S., & Hardy, I. (2017). Tracking the topological: The effects of standardised data upon teachers' practice. *British Journal of Educational Studies*, *65*(2), 219–238.

Lingard, B. (2009). Testing times: The need for new intelligent accountabilities for schooling. *QTU Professional Magazine*, *24*(November), 13–19.

Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education*, *51*(2), 129–147.

Lingard, B., Baroutsis, A., & Sellar, S. (2014). *Learning Commission report: Connecting schools with communities*. Brisbane, Qld: The University of Queensland.

Lingard, B., Creagh, S., & Vass, G. (2016). Education policy as numbers: Data categories and two Australian cases of misrecognition. *Journal of Education Policy*, *27*(3), 315–333.

Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Education Policy, 28*(5), 539–556.

Lingard, B., & Rawolle. S. (2004). Mediatizing educational policy: The journalistic field, science policy, and cross-field effects. *Journal of Education Policy*, *19*(3), 353–372.

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., Sellar, S., & Lewis, S. (2017, July). Accountabilities in schools and school systems. In G. W. Noblit (Ed.), *Oxford Research Encyclopedia of Education*. New York, NY: Oxford University Press. doi: 10.1093/acrefore/9780190264093.013.74

Lingard, B., Sellar, S., & Savage, G. (2014). Rearticulating social justice as equity in schooling policy: The effects of testing and data infrastructures. *British Journal of Sociology of Education*, *35*, 710–730.

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.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(4), 4–16.

Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy education in Australia? *Literacy Learning: The Middle Years*, *19*(2), 9.

Looney, A., Cumming, J., van der Kleij, F., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*. DOI: 0.1080/0969594X.2016.1268090

Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing*. Charlotte, NC: IAP.

Marks, G. (2014). Demographic and socioeconomic inequalities in student achievement over the school career. *Australian Journal of Education, 58*(3), 223–247. doi:10.1177/0004944114537052

Masters, G. N. (2009a). *Improving literacy, numeracy and science learning in Queensland primary schools: Preliminary advice*. Melbourne, Australia: Australian Council for Educational

Masters, G. N. (2009b). *Improving literacy, numeracy and science learning in Queensland primary schools*. Melbourne, Australia: Australian Council for Educational Research.

Matters, G. (2018). *Queensland NAPLAN Review: Parent perceptions report*. Brisbane: Queensland Department of Education.

Maxwell, G. S. (in preparation*). Improving student learning through use of data: Perspectives, practices and possibilities.* Dordrecht, The Netherlands: Springer.

Mayes, E., & Howell, A. (2018). The (hidden) injuries of NAPLAN: Two standardised test events and the making of 'at risk' student subjects. *International Journal of Inclusive Education*, 1–16.

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (1989). *The Hobart Declaration on schooling.* http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Hobart-Declaration-on-Schooling-1989.aspx

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (1997a). *Information statement 6th MCEETYA meeting, Melbourne, 14 March 1997*. Retrieved from http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiques%20 and%20Media%20Releases/Previous%20Council%20info%20statements/MCEETYA%20meeting% 20info%20statements/MC06_information_statement.pdf

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (1997b). *Information statement 8th MCEETYA meeting*, Melbourne, 11 December 1997. Retrieved from

http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Communiques%20
and%20Media%20Releases/Previous%20Council%20info%20statements/MCEETYA%20meeting%
20info%20statements/MC08_information_statement.pdf

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (1999). *The Adelaide Declaration on national goals for schooling in the twenty-first century*. Retrieved from http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Adelaide-Declaration.aspx

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (2007). *National report on schooling. Preliminary paper.*
http://www.curriculum.edu.au/verve/_resources/anr2007bmrks-layout_final.pdf

Ministerial Council for Education, Employment, Training and Youth Affairs [MCEETYA]. (2008). *The Melbourne Declaration on educational goals for young Australians*. Melbourne: MCEETYA. Retrieved from
http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Reports%20and%2
0publications/Publications/National%20goals%20for%20schooling/National_Declaration_on_the
_Educational_Goals_for_Young_Australians.pdf

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.

Mockler, N. (2016). NAPLAN and the problem frame: Exploring representations of NAPLAN in the print media, 2010 and 2013. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 181–198). London, England: Routledge.

Morley, P. (2011). Victorian Indigenous children's responses to Mathematics NAPLAN items. In *Mathematics: Traditions and new practices (Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia)* (pp. 523–530).

National Assessment of Educational Progress [NAEP]. (n.d.). *The Nation's report card. NAEP long-term trend assessments 2012. Summary of major findings*. Retrieved from
https://www.nationsreportcard.gov/ltt_2012/summary.aspx

Ng, C., Wyatt-Smit, C., & Bartlett, B. (2016). Disadvantaged students' voices on national testing: The submersion of NAPLAN's formative potential. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 152–163). London, England: Routledge.

No Child Left Behind (NCLB) Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Norton, S. (2009). The responses of one school to the 2008 Year 9 NAPLAN numeracy test. *Australian Mathematics Teacher*, *65*(4), 26–37.

O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review, 72*(3), 293–329.

Organisation for Economic Co-Operation and Development [OECD]. (2006). *PISA 2006: Science competencies for tomorrow's world.* Paris, France: OECD. Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/pisa 2006results.htm#ES

Organisation for Economic Co-Operation and Development [OECD]. (2008). *Education at a glance 2008*. Paris: OECD. https://www.oecd.org/education/skills-beyond-school/41284038.pdf

Perso, T. (2009). Cracking the NAPLAN code: Numeracy and literacy demands. *Australian Primary Mathematics Classroom*, *14*(3), 14–18.

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: The Belknap Press of Harvard University Press.

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.

Pierce, R., & Chick, H. L. (2011). Teachers' intentions to use national literacy and numeracy assessment data: A pilot study. *Australian Educational Researcher*, *38*, 433–447.

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.

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

Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy*, *29*(5), 640–657.

Power, S., & Frandji, D. (2010). Education markets, the new politics of recognition and the increasing fatalism towards inequality. *Journal of Education Policy, 25*, 385–396.

Queensland Teachers Union [QTU]. (2018). *QTU member survey on NAPLAN and MySchool*. Brisbane, Australia: Queensland Teachers Union.

Quinnell, L., & Carter, L. (2011). Cracking the language code: NAPLAN numeracy tests in Years 7 and 9. *Literacy Learning: The Middle Years*, *19*(1), 45–53.

Ragusa, A. T., & Bousfield, K. (2017). 'It's not the test, it's how it's used!' Critical analysis of public response to NAPLAN and My School senate inquiry. *British Journal of Sociology of Education*, *38*(3), 265–286.

Ranson, S. (2003). Public accountability in the age of neo-liberal governance. *Journal of Education Policy, 18*, 459–480.

Renshaw, P., Baroutsis, A., van Kraayenoord, C., Goos, M., & Dole, S. (2013). *Teachers using classroom data well: Identifying key features of effective practices. Final report.* Brisbane, Australia: The University of Queensland.

Rice, S., Dulfer, N., Polesel, J., & O'Hanlon, C. (2016). NAPLAN and student wellbeing: Teacher perceptions of the impact of NAPLAN on students. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 72–85). London, England: Routledge.

Rogers, S., Barblett, L., & Robinson, K. (2016). *Investigating the impact of NAPLAN on student and teacher emotional distress in independent schools*. Perth, Australia: Edith Cowan University.

Rogers, S. L., Barblett, L., & Robinson, K. (2018). Parent and teacher perceptions of NAPLAN in a sample of Independent schools in Western Australia. *Australian Educational Researcher*, *45*, 493–513.

Russell, M., Madaus, G., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: IAP.

Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, *11*, 45–61.

Sahlberg, P. (2011). *Finnish Lessons: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.

Salmon-Cox, L. (1981). Teachers and standardised tests: What's really happening? *Phi Delta Kappan, 6*(9), 631–634.

Senate Standing Committee on Education and Employment [SSCEE]. (2014). *Effectiveness of the National Assessment Program – Literacy and Numeracy. Final report.* Canberra, ACT: SSCEE.

Sharratt, L., & Fullan, M. (2012). *Putting faces on the data: What great leaders do!* Thousand Oaks, CA: Corwin & Ontario Principals' Council.

Shepard, L. A. (2003). Reconsidering large-scale assessment to heighten its relevance to learning. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

Singh, P., Märtsin, M., & Glasswell, K. (2015). Dilemmatic spaces: High-stakes testing and the possibilities of collaborative knowledge work to generate learning innovations. *Teachers and Teaching*, *21*(4), 379–399.

Stobart, G., & Eggen, T. (2012). High-stakes testing–value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, *19*(1), 1–6.

Spielman, A. (2017). *HMCI's commentary: Recent primary and secondary curriculum research*. London, England: Ofsted [online]. Retrieved from https://www.gov.uk/government/speeches/hmcis-commentary-october-2017

Standards and Testing Agency. (2018a). *2018/19 teacher assessment frameworks at the end of key stage 1. For use from the 2018/19 academic year onwards*. Retrieved from

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/740343/2018-19_teacher_assessment_frameworks_at_the_end_of_key_stage_1_WEBHO.pdf

Standards and Testing Agency [STA]. (2018b). *Pre-key stage 2 standards.* Retrieved from https://www.gov.uk/government/publications/pre-key-stage-2-standards?utm_source=60368303-949b-41e5-8565-80d28148e66b&utm_medium=email&utm_campaign=govuk-notifications&utm_content=immediate

Standards and Testing Agency [STA]. (n.d). *Guidance. Key stage 1 and key stage 2 test dates.* Retrieved from https://www.gov.uk/guidance/key-stage-1-and-key-stage-2-test-dates

Stecher, B. M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment*, *7*, 259–281.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London, England: Routledge.

Stobart, G., & Eggen, T. (2012). High-stakes testing–value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, *19*, 1–6. DOI: 10.1080/0969594X.2012.639191

Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison Project School. *American Journal of Evaluation, 25*, 277–293.

Swain, K., Pendergast, D., & Cumming, J. (2018). Student experiences of NAPLAN: Sharing insights from two school sites. *The Australian Educational Researcher*, *45*(3), 315–342.

Takayama, K., & Lingard, B. (2018, advance online publication). Datafication of schooling in Japan: An epistemic critique through the 'problem of Japanese education'. *Journal of Education Policy,* 1–21. doi:10.1080/02680939.208.1518542

Tan, C. (2019). *Comparing high-performing education systems: Understanding Singapore, Shanghai, and Hong Kong*. London, England: Routledge.

The Sunday Mail. (2018, September 16). *Ultimate schools guide.* Brisbane, Australia: The Sunday Mail.

Thompson, G. (2012). *Effects of NAPLAN: Executive summary*. Murdoch, Australia: Murdoch University. Retrieved from http://www.literacyeducators.com.au/wp-content/uploads/2013/12/Executive-Summary.pdf

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. (2016). Local experiences, global similarities: Teacher perceptions of the impacts of national testing. In B. Lingard, G. Thompson & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 57–71). London, England: Routledge.

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

Thompson, G., & Cook, I. (2014). Manipulating the data: Teaching and NAPLAN in the control society. *Discourse: Studies in the Cultural Politics of Education*, *35*(1), 129–142.

Thompson, G., & Harbaugh, A.G. (2012). *The effects of NAPLAN: Teacher perceptions of the impact on pedagogy and curriculum*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Sydney. https://eprints.qut.edu.au/86167/1/86167.pdf

Thompson, G., & Harbaugh, A. G. (2013). A preliminary analysis of teacher perceptions of the effects of NAPLAN on pedagogy and curriculum. *Australian Educational Researcher, 40*, 299–314. doi:10.1007/s13384-013-0093-0

Thompson, G., & Mockler, N. (2016). Principals of audit: Testing, data and 'implicated advocacy'. *Journal of Educational Administration and History*, *48*(1), 1–18.

Thompson, G., Sellar, S. & Lingard, B. (2016). The life of data: Evolving national testing. In B. Lingard, G. Thompson & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 212–229). London, England: Routledge.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal*, *42*(3), 454–476.

Vass, G. & Chalmers, G. (2016). NAPLAN, achievement gaps and embedding indigenous perspectives in schooling. In B. Lingard, G. Thompson & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 139–151). London, England: Routledge.

Wahlstrom, K.L., Seashore Lewis K., Leith wood, K., & Anderson, S.E. (2010). *Investigating the links to improved student learning: Executive summary of research findings*. (Learning from Leadership Project). St. Paul, MN: Center for Applied Research and Educational Improvement (CAREI), University of Minnesota.

Ward, D. M. (2012). *The effects of standardised assessment (NAPLAN) on teacher pedagogy at two Queensland schools* (Doctoral dissertation, Queensland University of Technology).

Watt, G., Finger, G., Smart, V., & Banjer, F. (2014, September). Project 600: Inspire, connect and transform. In *26th Australian Computers in Education Conference, Adelaide.* http://acec2014.acce.edu.au/sites/2014/files/attachments/ACEC2014_Project%20600%20Inspire%20Connect%20and%20Transform_Final_Watt_Finger_Smart_Banjer.pdf

Whitlam Institute. (2013). *The experience of education: The impacts of high stakes testing on school students and their families (Parental attitudes and perceptions concerning NAPLAN)*. Sydney, Australia: Whitlam Institute, University of Western Sydney.

Wigglesworth, G., Simpson, J., & Loakes, D. (2011). NAPLAN language assessments for Indigenous children in remote communities: Issues and problems. *Australian Review of Applied Linguistics*, *34*(3), 320–343.

Wu, M. (2011). The use of NAPLAN data for English language teaching. *Idiom*, *47*(1), 38.

Wu, M. (2016). What national testing data can tell us. In B. Lingard, G. Thompson & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 18–29). London, England: Routledge.

Wyatt-Smith, C. M., Adie, L., & Harris, L. (2018, June). Data walls and classroom learning. *Newsmonth*. Retrieved from http://publications.ieu.asn.au/2018-june-newsmonth/feature/data-walls-and-classroom-learning/

Wyatt-Smith, C., Bridges, S., Hedemann, M., & Neville, M. (2008). Designing professional learning for effecting change: Partnerships for local and system networks. *The Australian Educational Researcher, 35*(3), 1–20.

Wyatt-Smith, C. M., Harris, L., & Adie, L. (2018, August). Data walls and evidence of impact. *Newsmonth.* Retrieved from http://publications.ieu.asn.au/newsmonth-29/feature1/examining-evidence-impact/

Wyatt-Smith, C., & Jackson, C. (2016). NAPLAN data on writing: A picture of accelerating negative change. *Australian Journal of Language and Literacy, 39*(3), 233–244*.*

Wyatt-Smith, C., Jackson, C., Adie, L., Humphrey, S., Wang, J., & Hollis, C. (2017). *Research partnerships and improvement science: Using data to inform the teaching of writing and assessment*. Final report. Brisbane, Australia: Institute for Learning Sciences and Teacher Education. ISBN: 978-1-922097-52-1

Wyn, J., Turnbull, M., & Grimshaw, L. (2014). *The experience of education: The impacts of high stakes testing on school students and families (A qualitative study)*. Sydney, Australia: The Whitlam Institute, University of Western Sydney.

Zammit, K. (2018, May 15). Re-envisioning NAPLAN: Use technology to make the tests more authentic and relevant. *The Conversation*. Retrieved from https://theconversation.com/re-envisioning-naplan-use-technology-to-make-the-tests-more-authentic-and-relevant-95035