Cumming, J., Hand, K., & Goldstein, H. (2020). Enhanced use of educational accountability data to monitor educational progress of Australian students with focus on Indigenous students. *Educational Assessment, Evaluation and Accountability* , *32*, 29-51. https://doi.org/10.1007/s11092-019-09310-x

Enhanced use of educational accountability data to monitor educational progress of Australian students with focus on Indigenous students.

In: Educational Assessment, Evaluation and Accountability (To appear)

Joy Cumming *
Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Australia 4000
Orcid id: 0000-0001-9221-7677

Harvey Goldstein
Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Australia 4000
and
School of Education, University of Bristol, Bristol, BS8 1JA, UK
Orcid id: 0000-0003-3878-4122

Kirstine Hand
Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Australia 4000 (at time of contribution)
Orcid id: 0000-0002-0566-4434

* Corresponding Author
Contact details
email: joy.cumming@acu.edu.au
telephone: +61 7 3623 7862
Postal Address: ILSTE, GPO Box 2587, Brisbane, QLD, 4001, Australia

[Short Title] Enhanced longitudinal analyses Australian NAPLAN data

Abstract

In Australia, under the National Assessment Plan, educational accountability testing in literacy and numeracy (NAPLAN) is undertaken with all students in Years 3, 5, 7 and 9 to monitor student achievement and inform policy. However, the extent to which these data

have been analyzed to report student progress is limited. This article reports a study analyzing Year 3 and Year 5 NAPLAN reading and numeracy data, school and student information for a single student cohort from Queensland, Australia, to examine student achievement and progress. The analyses use longitudinal multilevel modelling, incorporating an enhanced approach for missing data imputation, given that such data frequently involve large amounts of missing data and failure to account properly for such missing data may bias interpretations of analyses. Further, statistical adjustments to deal with the impact of measurement error, an aspect not previously addressed in such analyses of data, are undertaken. An especial focus of analyses is achievement of Australian Indigenous and non-Indigenous students. International and national data demonstrate a considerable achievement gap between these students. "Closing the gap" is a core Australian education equity policy, with NAPLAN data used as a primary indicator of policy impact. Overall, analyses indicate greater understanding of student progress for all students is available from Australian data if appropriate analyses are undertaken. However, analyses also demonstrate not only that the gap between Australian Indigenous and non-Indigenous student progress increases as they move through school but also diversity of achievement within the Indigenous student cohort. Implications for policy are considered.

Keywords: educational accountability; longitudinal multilevel modelling; missing data; measurement error; Australian Indigenous students

Introduction

This article reports a study examining Australian accountability data from the National Assessment Program—Literacy and Numeracy (NAPLAN), common national literacy and numeracy tests introduced in 2008, using improved methods of analyzing these existing and comprehensive data. The starting point for the study was our perception that current analyses and public reporting of NAPLAN data are insufficient to realize the potential of the data to monitor student progress and inform policy. Therefore, the first aim of the study was to use appropriate statistical techniques with NAPLAN data and related demographic student and school characteristics to improve the current level of monitoring student learning improvement over time.

NAPLAN tests measure four domains: Reading; Writing; Language Conventions (Spelling, Grammar and Punctuation); and Numeracy. Tests are designed to describe achievement and progression from Year 3 to Year 9 using a ten-band scale, with six overlapping bands assigned to each Year level (ACARA 2016a). One band at each Year level represents a national minimum standard or benchmark that all students in that Year are 'expected' to meet. Student raw scores across the ten bands are scaled to the range 0 to 1000 (ACARA 2017). Participation in NAPLAN is mandatory for all students[1] in Years 3, 5, 7 and 9[2] in all Australian schools (government and non-government[3]) as a condition of public funding legislation and federal, state and territory[4] agreements. NAPLAN data therefore

[1] Students who have significant disability or are unable to 'access' the NAPLAN tests may be exempted. Students who are ill are noted as absent. Students may be withdrawn if parents or students do not wish to participate. Reported statistics (means and percentages) include statistical imputation for absent or withdrawn students; exempt students are not included in the population or calculation of results (ACARA 2017). In 2013 for Year 5 Reading, Queensland, participation rate was 95.5% of students, with 2.3% absent, 2.7% withdrawn, and a 1.7% exempt (ACARA 2013), with similar rates for Numeracy.

[2] Australian school years are referred to as Year levels, rather than Grades, as students progress through schooling with their age cohort, rather than a 'grade' level of performance.

[3] Approximately two-thirds of Australian students attend government schools and one-third attend non-government schools (20% Catholic schools, 10% other independent schools).

[4] Australia is a federation of six states (New South Wales, Queensland, South Australia, Tasmania, Victoria, Western Australia) and two territories (Australian Capital Territory, Northern Territory).

provide a potentially substantial body of information on performance of the Australian student cohort each year, across Year levels and across time.

Our analyses of NAPLAN data use longitudinal multilevel analyses and incorporate developments in imputation for missing data, a concern in longitudinal NAPLAN data, and adjustments for measurement (un)reliability in model estimation. These approaches are applicable in many international contexts where educational accountability data are used to compare cohort performance and inform policy. The paper examines achievement and progress for all students and their import for policy and practice, , emulating comparisons of subgroups currently undertaken in national reporting (ACARA 2017).

The second aim and an especial focus of the analyses presented in this paper was to enhance current information on the academic progress of Indigenous students, as identified by NAPLAN data. The educational achievement of Australian Indigenous students has long been a major policy concern identified in 2008 as "a key priority" for the next decade: "Australia has failed to improve educational outcomes for many Indigenous Australians" (Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA] 2008, p. 5). This priority is referred to as the "Closing the Gap agenda" (Council of Australian Governments [COAG] 2009, p. A-16) with the goal to ensure the achievements of Indigenous students improve to "match those of other students" (MCEETYA 2008, p. 7). Indigenous students for NAPLAN data are defined as those who identify "as being of Aboriginal and/or of Torres Strait Islander origin" where "origin" relates to "Australian Aboriginal or Torres Strait Islander descent and for some, but not all, their cultural identity" (Australian Curriculum, Assessment and Reporting Authority [ACARA] 2017, p. vii). 'Aboriginal and Torres Strait Islander' are collective terms reflecting both heritage and conceptual identity and encompassing many peoples with different languages and identification with country and place. 'Cultural identity' goes beyond biological descent. In

some contexts, to identify oneself as Indigenous may also require being accepted by an Indigenous community (Australian Government, Australian Law Reform Commission 2010). However, NAPLAN records rely on self-identification in school enrolment data. In 2013, relevant to the data in this study, over five per cent (183,000) of Australian students identified as Indigenous (Aboriginal and Torres Strait Islander) (Australian Bureau of Statistics [ABS] 2014).

Differences in achievement between Australian Indigenous and non-Indigenous students are consistently demonstrated in numerous international test programs. In PISA 2015 Scientific Literacy, Indigenous students performed at a lower level than non-Indigenous Australian students, equated to $2^{1}/_{2}$ years of schooling, and were lower on average than the OECD average, whereas non-Indigenous students on average were above the OECD average (Thomson, De Bortoli, and Underwood 2017). Indigenous students were underrepresented in high proficiency bands and overrepresented in the lowest proficiency bands, compared with non-Indigenous students. Similar outcomes occurred for PISA Reading Literacy and Mathematical Literacy, and for TIMSS (Trends in International Mathematics and Science Study) (Years 4 and 8) (Thomson, Wernert et al. 2017) and PIRLS (Progress in International Reading Literacy Study) (Year 4) (Thomson, Hillman, Schmid, Rodrigues, and Fullarton 2017). Indigenous students have considerably lower achievement than their non-Indigenous peers, the gap has not closed and indeed has increased over two cycles of PIRLS (2011, 2016), more Indigenous students demonstrate low levels of achievement, and few Indigenous students achieve at high levels, when compared with their non-Indigenous peers.

The first Australian goal espoused in the Melbourne Declaration on Educational Goals for Young Australians (MCEETYA 2008) is promotion of equity and excellence for all students, including removal of discrimination and disadvantage based on “gender, language, … ethnicity, … socioeconomic background or geographic location” (p. 7). The national

assessment program is identified as providing "education ministers with information about the success of their policies and resourcing in priority curriculum areas … [and] capacity to monitor the success of policies aimed at improving the achievement of different student groups, such as Indigenous students" (ACARA 2016c).

As a result, annual national reports on NAPLAN achievement compare student achievement on NAPLAN for equity subgroups identified in the Melbourne Declaration, that is, for gender, language background (English/not English), ethnicity (Indigenous/not Indigenous), socio-economic background (parental education and occupation) and geographic location. NAPLAN is identified as providing "robust data to inform and support improvements to teaching and learning practices in Australian schools" (ACARA 2013, p. iv), although policy implications are not discussed. NAPLAN data are key policy indicators of the educational achievement for Indigenous and non-Indigenous students.

Potential bias in NAPLAN

NAPLAN tests are criticized on various grounds: equity concerns; validity of interpretation for different purposes (Thompson, Adie, and Klenowski 2017); and impact on schools and teachers (Lingard, Thompson, and Sellar 2016), and students (Author; Rice, Dulfer, Polesol, and O'Hanlon 2016). Equity concerns for measurement of achievement for Indigenous students are language and cultural bias (Klenowski 2009; Klenowski, Tobias, Funnell, Vance, and Kaesehagen 2010; Warren, Young, and de Vries 2007).

NAPLAN test forms give primacy to language, and, more specifically, to standard Australian English, potentially affecting students who do not have English as a first language. Many Indigenous students in remote or very remote areas speak an Aboriginal language or Aboriginal English as their first language. Equity statements in policies for all students identify that "schools [should] build on local cultural knowledge and experience of Indigenous students as a foundation for learning" (MCEETYA 2008, p. 7).

However, Australian education policy identifies standard Australian English as an expected literacy outcome for all students to achieve equity education goals:

> While it is recognized that all students do not have the same cultural and language background, competency in Australian English is essential for all students to allow them to participate fully in Australian society. This is particularly important for students from Aboriginal and Torres Strait Island heritages who may not speak English as their first, second or third language… (Australian Government 2014, p. 6)

NAPLAN tests are based on Australian Curriculum content, with representatives for Indigenous and non-Indigenous students, states and territories, and geo-locations engaged in preliminary trialling and examination of items and responses for potential bias (ACARA 2018). Psychometric procedures are used to determine cultural fairness. Technical reports provided for NAPLAN indicate that differential item functioning analyses are undertaken to explore potential discriminatory test items (DIF) with different subgroups, including gender, language background and Indigenous culture (ACARA 2014). Subgroup performance on individual items is examined in terms of relative difficulty for comparative groups, that is, for example, Indigenous and non-Indigenous students. However, the 2013 ACARA NAPLAN Technical Report notes that 'there may not be many Indigenous students along parts of the ability range' (ACARA 2014, p. 35). In the NAPLAN tests used in this study, some 30 per cent of Year 5 Reading items were indicated after administration as having different relative difficulty for Indigenous students compared with non-Indigenous students and 22.5 per cent Year 5 Numeracy items similarly. These items may be either relatively easier or more difficult for Indigenous students than other students. As these results are obtained after test administration, no adjustments were made to the NAPLAN tests to account for these differences. No Technical Report for the 2011 NAPLAN has been made public.

NAPLAN therefore undertakes standard psychometric procedures to prevent bias in test development. Even studies taking a sociological perspective to examine culture-fair

assessment and teaching within classrooms (see, e.g., Klenowski and Gertz 2009) have used NAPLAN data as one measure of learning outcomes and improvement for students. Klenowski and Gertz noted that differences in such outcomes for Indigenous students were not necessarily due to test bias but may be due to limited student experiences with the content and mode of the tests, demonstrating the need for professional development in strategies that can address these factors, including the need to focus on language (Klenowski et al. 2010).

In a recent comprehensive review of NAPLAN in Queensland (Cumming et al. 2018), several teachers identified NAPLAN testing as presenting language barriers for Indigenous students, especially in the early years, and culturally-biased. However, Indigenous representatives, when consulted, considered that, despite the relatively poor performance overall of Indigenous students, the monitoring of achievement and progress of Indigenous students provided by NAPLAN was a "valuable tool, even if it is highlighting [the] elephant in the room" (p. 114). Such monitoring was seen to give Indigenous students and families a "voice". Parents of students who identified as Indigenous were reported to be more likely to "ensure their child's participation" and "valued the NAPLAN process to identify their child's achievement" (p. 128). In the 2018 Queensland review, Indigenous sector representatives and parents aspired to the same outcomes for Indigenous students as for non-Indigenous students, emphasizing the need for high expectations comparable to those for other students, not just minimum standards that are the focus of 'closing the gap' policies (Cumming et al., 2018, p. 114).

Concerns about NAPLAN led to a federal Senate inquiry (Senate Standing Committee on Education and Employment 2014). While noting the range of issues about the nature and impact of NAPLAN expressed to the inquiry, including language and cultural suitability for Indigenous students, the Senate Committee accepted the perspective that NAPLAN data are

of “significant value to…students, schools, parents, education authorities, the wider community, and state and national governments” (p. 25).

NAPLAN is thus generally described as providing a nationally-comparable snapshot of student achievement. Despite potential flaws in NAPLAN for students from different cultural backgrounds and disadvantage, NAPLAN, aligned with the Australian Curriculum and expectations for all students to achieve the same goals and demonstrate achievement in similar ways, has become the benchmark for student accountability, learning and progress. The uses made of NAPLAN test outcomes and other systematically collected data to judge improvement and student achievement, therefore, emphasizes the need for use of appropriate methodologies and analyses to inform future policy and strategic directions in order to improve student learning outcomes.

Australian Policy, Goals and Identified Achievement for Indigenous Student Education

National policy documents and strategies to address Indigenous education have been in place for several decades, with various Action Plans developed to “accelerate” Indigenous student progress in literacy and numeracy (MCEETYA 2006, p. 4). As Australian Indigenous peoples are identified as at risk not only in terms of educational failure and underachievement but in many significant areas such as infant mortality, health and access to services, and employment, a *National Integrated Strategy for Closing the Gap in Indigenous Disadvantage* (COAG 2009) was developed to address all aspects of disadvantage. It incorporated a *National Indigenous Reform Agreement*, revisited annually. The Agreement identifies “gaps”, targets, and progress against targets. For education, the progress indicator is the proportion of Indigenous students at or above the national minimum literacy and numeracy standards, as measured by NAPLAN. The initial agreement established a target to halve the gaps, that is, halve the difference in the respective proportions of Indigenous and non-Indigenous students achieving minimum standards in literacy and numeracy, by 2018. A later agreement set more

ambitious goals, that is, to *more* than halve the gap by 2025 (COAG 2013). The Indigenous Reform Agreements unite several strategies to improve Indigenous student achievement (Ministerial Council for Education, Early Childhood Development and Youth Affairs [MCEECDYA] 2009),[5] including considerable financial resources allocated to selected schools across Australia, loadings for additional support for Aboriginal and Torres Strait Islander students in all schools, and for literacy projects in remote schools.

As noted, NAPLAN data on achievement and progress of Indigenous Australian students play a significant role as targets and indicators to evaluate the success of these strategies. The issue addressed in this paper is that, for all equity variables identified in policy as of interest, analyses of NAPLAN data typically occur at a simplistic level. Annual national reports provide analyses of cross-sectional data of average NAPLAN achievement for each state and territory, and for student subgroups, including gender, Indigenous status, home language (English, other than English), geo-location, parental education and parental occupation (see, e.g., ACARA 2017). Few interactions are examined, although achievement of Indigenous students by geo-location is reported. Longitudinal time series analyses of NAPLAN data are presented but based on average scores not tracking of individual achievement.

Based on these analyses, national NAPLAN reports demonstrate the Indigenous–non-Indigenous achievement gap, measured in terms of year progress equivalents, to be an approximate but consistent two-year lag, that is, Year 5 Indigenous students perform at approximately the same level as Year 3 non-Indigenous students, and Year 7 and Year 9

[5] For example, the Action Plan (2010-2014) stated: "Reform directions are detailed in the National Indigenous Reform Agreement, the National Education Agreement, the Early Childhood Education National Partnership, the Indigenous Early Childhood Development National Partnership, Remote Service Delivery National Partnership, the Smarter Schools - Improving Teacher Quality National Partnership, the Smarter Schools - Low Socio-economic Status School Communities National Partnership and the Smarter Schools - Literacy and Numeracy National Partnership, the Youth Achievement and Transitions National Partnership and other agreements. Commitments made in these national partnerships and agreements are brought together in the Plan with a number of new and continuing complementary measures to close the gap between the educational outcomes of Aboriginal and Torres Strait Islander students and other students." (MCEECDYA 2009, p. 4)

Indigenous students perform at similar, but slightly lower levels, to Year 5 and Year 7 non-Indigenous students, respectively. These results apply across all domains tested (see, e.g., ACARA 2017). Longitudinal analyses demonstrate slightly improved average outcomes for both Indigenous and non-Indigenous students overall, from the baseline data of 2008 to 2015, but no evidence of Indigenous students 'closing the gap' on their non-Indigenous peers in terms of progress over this period. Based on cross-sectional, not longitudinal analyses, and despite funding initiatives and strategies, the initial 2009 target to halve the gap between Indigenous and non-Indigenous students in literacy and numeracy by 2018 is recognized as simply "not on track" (Commonwealth of Australia, Department of the Prime Minister and Cabinet 2018, p. 58).

A factor related to Indigenous student achievement in NAPLAN is 'geo-location' (see Online Supplement Section 1). While approximately 34 per cent of the Indigenous population live in Australia's major cities and 44 per cent in inner and outer regional areas, more than 20 per cent live in remote and very remote areas (ABS 2013b). Based on cross-sectional data, the gap in literacy and numeracy achievement for Indigenous students in remote and very remote areas is "amplified" (Guenther, Bar, and Osborne 2013, p. 101) when compared with the gap for Indigenous students in metropolitan and provincial locations. This pattern does not occur for non-Indigenous students (ACARA 2017; ACIL Allen Consulting with La Trobe University and Phillips KPA [ACIL] 2014; Steering Committee for the Review of Government Service Provision 2014). Location may interact with language. In Queensland, in the dataset analyzed for the study presented in this article, 10.5 per cent of Indigenous students in Metropolitan and 8.7 per cent in Provincial regions indicated a language background other than English, that is, a language other than English was spoken in the home. The proportion increased to 49 per cent for Indigenous students in Remote and Very Remote regions. However, such interactions are difficult to explore statistically due to the

small number of students involved. For the Queensland percentages noted above, 49 per cent of Indigenous students in Remote and Very Remote regions equates to only 317 students of the more than 53,000 students in the cohort, and for the 10.5 and 8.7 percentages for metropolitan and provincial regions, 189 and 112 students respectively.

Research Purpose

The Australian NAPLAN accountability data not only collect test outcomes for full cohorts of students in Years 3, 5, 7 and 9 but also record individual student outcomes longitudinally through use of school and student identification codes. The starting point for the present study is that current national reporting of NAPLAN student outcomes, using limited statistical analyses and representations, fails to fulfil the potential of these longitudinal data to inform policy and capitalize on the data to inform future strategic directions. In this article we report innovative analyses undertaken to examine the achievements and progress of students from Year 3 to Year 5 and the impact of student gender, home language and geo-location as variables used in the national reports of schooling to compare student performance, reflecting the equity groups identified in education policy goals. Interactions between explanatory variables are examined. The analyses use two statistical developments for educational accountability data analyses. First, an enhanced process for imputation (Goldstein, Carpenter, and Browne 2014) is used to address attrition in longitudinal data sets. The second innovation addresses the effects on parameter estimates and substantive inferences when account is taken of measurement errors, especially in covariates. Goldstein, Browne and Charlton (2018) have shown how to take account of missing data and measurement errors within a single model and this procedure is utilized in the present paper. It is well known that large measurement errors (low measurement reliabilities) can substantially alter estimates (see, e.g., Goldstein and French 2015), but typically little attempt is made to adjust for these in analyses.

As noted, a particular focus is analyses of NAPLAN data and student growth from Year 3 to Year 5 for Indigenous students in comparison with non-Indigenous students. The analyses therefore use student identification as Indigenous as a further variable in analyses, in combination with the other variables. Unfortunately, as noted, the number of students who are Indigenous and indicate a language background other than English was too small for meaningful interactions to be undertaken.

The research questions therefore are the extent to which more statistically-appropriate analyses of NAPLAN provide information on student achievement and progress across two year levels, and more specifically, the extent to which these analyses of NAPLAN data provide greater enlightenment regarding a key Australian policy focus, achievement and progress of Indigenous students, and whether indeed the "gap" is closing.

Data and Sample Characteristics

Since 2008, NAPLAN provides large-scale longitudinal data, linked for the majority of students, that allow us to map progress. The dataset used in the study was provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA), which administers NAPLAN at the national level. Although ACARA collects NAPLAN data for all Australian students, under agreements with each state and territory, ACARA's data sets *only* link data for students for each two consecutive years of testing, that is, Years 3 to 5, Years 5 to 7, and Years 7 to 9.

Data in the file provided by ACARA were de-identified but each record was provided with an individual ID and school ID. This study used data for the cohort of Queensland students in the first two years of NAPLAN testing, Years 3 to 5. The students participated in NAPLAN in Year 3 in 2011 and in Year 5 in 2013, with the majority of student records linked, as noted later. These students completed Year 9 NAPLAN in 2017 and will not

complete high school until 2020. The data therefore focus on the early years of schooling and achievement and potential implications for further schooling.

Our analyses focus on two NAPLAN outcome measures, Reading and Numeracy. The file provided by ACARA contained 53,054 cases with Year 5 NAPLAN scores in Reading and Numeracy. Sixteen student records that did not provide school identification codes were removed, leaving a data set with 53,038 students. Of these records, 47,506 students had participated in the Year 3 NAPLAN Reading test, with 5,532 or 10.4 per cent of students missing these scores, and 47,292 students had participated in the Year 3 NAPLAN Numeracy test, with 10.8 per cent of students missing these scores. Students who participated in both Year 3 tests numbered 47,060, 446 students had participated in Year 3 Reading but not Numeracy, and 232 students had participated in Year 3 Numeracy but not Reading. The data set included item responses and scaled scores. Individual student records provided by ACARA included data for the demographic variables gender, ethnic origin, language background, geo-location, and parental occupation and education level (self-report) (Online Supplement Section 1).

In 2013 (and continuing), Queensland, the third most populous state,[6] had the second largest proportion of Australia's Indigenous students (29%), exceeded only by New South Wales (32.0%), and the third highest proportion of Indigenous students within a system (7%), exceeded only by the Northern Territory (40%) and Tasmania (7%) (ABS 2014). It covers a range of geo-locations for schools attended by Indigenous students. Queensland, therefore, provides a strong educational context for examining educational achievement of Indigenous students and non-Indigenous students in the same schooling context.

Tables 1 and 2 provide descriptive statistics for NAPLAN Reading and Numeracy

[6] In 2013, relevant to the data analyzed, the proportion of the Australian population of 23 million in each state and territory was: NSW: 32%; Vic: 25%; Qld: 20%; WA: 11%; SA: 7%; Tas: 2%; ACT: 2%; NT: 1% (ABS, 2013a).

achievement for the population of Queensland students in Year 5 in the Year 2013 based on variables used in national reports.[7] We have no record of 2011 Year 3 students who were not present at Year 5. Table 1 shows that the proportions of female and male students are similar for both Indigenous and non-Indigenous students, that overall some seven per cent of Queensland students in Year 5 in 2013 identified as Indigenous, and over nine per cent of all students had a language other than English spoken in the home. Nearly all students lived in metropolitan and provincial regions, however, the proportion of Indigenous students living in remote areas (17.3%) was far greater than the proportion of non-Indigenous students (2.2%).

Insert Table 1 about here

Table 2 shows that overall over 5,000 students in Year 5 did not have NAPLAN data recorded for Year 3, with numbers varied slightly for Reading and Numeracy. Average Reading and Numeracy scaled scores show considerable changes for the overall population from Year 3 to Year 5. There is a slight reversal in performance for gender and Reading and Numeracy, although differences are not large. Indigenous students' average achievement for Reading and Numeracy for both Years 3 and 5 is noticeably worse than for non-Indigenous students. Language background does not appear to affect results; however regional effects are apparent with student achievement appearing to worsen as students' school location moves away from metropolitan areas.

Insert Table 2 about here

Variables

Variables incorporated in the modelling are elaborated in Online Supplement Section 1. Reading and Numeracy measures are detailed as well as the demographic variables included

[7] Between Years 3 and 5 approximately 20% of students change schools. Unfortunately, we have no information about when these changes take place so that accounting for exposure to two or, indeed, more schools is not possible and we do not pursue this further. Goldstein, Burgess and McConnell (2007) suggest that while a failure to take account of movement between schools affects the relative proportions of variance at school and student level, the effects on the remaining coefficients are small.

in the analyses reflecting equity groups of interest, that is, gender (female, male), Indigenous/non-Indigenous status, geo-location (metropolitan, provincial, remote/very remote with the last category combined to prevent school or student identification), and language background of students (English, other than English). Demographic data also record parental occupation and education as reported by parents on student enrolment at school. We do not use these variables in our analyses due to the extent of missing data in our dataset. A substantial proportion of the data, from 15 per cent for school education for mothers of non-Indigenous students up to 50 per cent for school education and over 50 per cent for occupation of fathers of Indigenous students, was missing. Parental education and occupation were more likely to be missing for parents of both non-Indigenous and Indigenous students in rural schools (see Tables S1 and S2 in the Online Supplement Section 2). No data for the remaining demographic variables in the data set of 53,038 Year 5 students were missing.

Methods

The inclusion of both school and student identification codes means that analyses were able to be undertaken using a hierarchical structure with students nested within their schools. We have accordingly fitted two-level models that incorporate specific school effects, a first variation from the overall analyses presented in national reports. In our analyses, we use normalized (mean=0, SD=1) test scores separately for each Year level so that effects are on comparable standard deviation scales. NAPLAN scaled scores have been subject to an equating procedure to constitute a scale providing 'comparable' scores across Year levels. While this form of 'vertical equating' is open to a number of criticisms (Goldstein and Wood 1989), these were not relevant for our analyses. More relevant are the distributional assumptions, namely normality, and this has motivated the choice of scale as above. In particular, we assume that the Cronbach α reliability estimates obtained from the Queensland

data for use in the longitudinal modelling are relatively unaffected by the scaling procedures adopted across the whole NAPLAN cohort.

Adjustment for Missing Data

The annual national NAPLAN reports incorporate conventional sampling errors[8] allowing, in principle, statistical comparisons. Although imputation procedures for missing responses are also used in scaling procedures, it is not clear in the national reports or accompanying technical reports (ACARA 2014) the extent to which analyses of average gain and progression adjust for missing student responses across successive years of testing, as numbers of students are not reported. Certainly, in the procedures used in the public reporting of individual school NAPLAN student progress across successive test Year levels (myschool.edu.au), growth is based on average scores for students who participated in tests in a school in both years. This procedure, often known as 'listwise deletion' in the statistical literature, is a common method for dealing with data that are missing. Where such missingness, however, is non-random this may introduce biases, both in terms of school comparisons and estimates of model parameters. Our first analyses, therefore, look at what these biases may be.

Cumming and Goldstein (2016) have shown that including information from predictor variables having missing data increases precision of parameter estimates. We apply the same Bayesian-based imputation model utilized by Cumming and Goldstein, with illustration as to how utilizing all available data can reduce biases. The following model for Reading illustrates the issue.

Table 2 demonstrated that, prima facie, no scaled mean differences at each Year level are obvious for students with and without English as their language background (e.g., Year 5,

[8] While NAPLAN includes full cohorts in testing, rather than sampling of students, NAPLAN technical reports note that sampling error is included at student and school level in order "to make inference about the educational systems each year and not about the specific student cohorts" within a year, and also to take account of "a certain amount of non-response" (ACARA 2016b).

mean scaled scores 489.33 for students with a language other than English spoken in the home, 499.01 for students with only English spoken in the home), when missing data are excluded. However, as Table 3 shows, those missing Year 3 data have quite different Year 5 mean scores from those of students not missing Year 3 data. Table 3 uses scaled scores converted to normalized NAPLAN scores distributed as noted above.

Insert Table 3 about here

Noticeable differences emerge in the mean scores for students in each subgroup according to whether they are missing or not missing the Year 3 score. Those missing Year 3 scores having significantly lower Year 5 test scores in all cases than those who have Year 3 scores present.

To illustrate the effect on longitudinal model estimates we fit the following model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + u_j + e_{ij} \quad (1)$$

$$u_j \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

with variables identified in Table 4. Table 4 contrasts estimates obtained from deleting student records with a missing Year 3 test score with those obtained from analysis that utilizes all available date through incorporation of the imputation model. Once again, Reading is used for demonstration, with the incorporated variables of Indigenous status, gender and language background, and the interaction between gender and Indigenous status.

Insert Table 4 about here

The comparison shows that the effect of Indigenous membership becomes more negative when imputation is undertaken. The more striking result is that the effect of a language background other than English, while effectively zero in the unadjusted analysis, becomes negative and statistically significant when adjustment is made for missing values. There is no apparent effect based on gender.

Adjusting for Measurement Error

In our final longitudinal models, we also make adjustments as noted for measurement error, a factor that needs to be considered when undertaking analyses of achievement data, that are not in themselves perfect measures. We adopt a simple measurement error model for an observed variable x_1 where, dropping subscripts,

$$x_1 = X_1 + \gamma_1 \quad (2)$$

$$Y = X\beta + e \quad (3)$$

Lower case letters denote observed values and upper case true values. We fit the true model of interest, (4) below, using a Bayesian MCMC algorithm. Online Supplement Section 3 describes the detailed algorithm steps and also shows how to adjust for measurement error in the response. Since the tests are two years apart we assume that measurement errors are independent. Online Supplement Section 4 provides reliability estimates using conventional 'split test' methods for Years 3 and 5 student data for Reading and Numeracy. The estimated overall reliabilities based on Cronbach's α are approximately 0.85, with a range from 0.83 to 0.88 for subgroups defined by gender and Indigenous status, and this value is used in the following analyses.

Results

The following analyses of our dataset present models of increasing complexity to demonstrate the extent to which enhanced modelling and analyses can interrogate student achievement, differences and progress. A primary strength of these models is the use of individual student data as opposed to the national report analyses using only average data. The first model analyzes both Year 3 Reading and Numeracy and Year 5 Reading and Numeracy outcomes separately. This is a 2 level model where students are treated as nested within their Year 3 or Year 5 school (Goldstein 2011). Table 5 shows this basic model incorporating Indigenous status (x_{1ij}, Indigenous=1, non-Indigenous=0) and gender

($x_{2ij,}$ female =1, male=0). We also include an interaction for these two variables. For Reading, 5,532 students have no Year 3 Reading scores and 5,766 students have no Numeracy scores. These are incorporated via imputation as described previously.

The generic model can be written as

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij} x_{2ij} + u_j + e_{ij} \quad (4)$$

$$u_j \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

Insert Table 5 about here

We see from this simple model similar patterns in both Year 3 and Year 5. Indigenous students do considerably worse than non-Indigenous students in both Reading and Numeracy. Girls do better than boys for Reading but the reverse is true for Numeracy. Of some interest is the interaction between gender and Indigenous status: for Reading if the student is Indigenous the gender difference is effectively the same as for non-Indigenous students since the interaction term is non-significant, whereas for Numeracy in Year 5, the female–male difference for Indigenous students is not as large as for non-Indigenous students. Current Australian national NAPLAN reports do not examine such straightforward interactions, even for these cross-sectional data. These results are explored further in later analyses.

The second multilevel model using the same variable set and interaction as the previous model encompasses longitudinal analyses to explore student progress in Reading and Numeracy between Year 3 and Year 5. The model is now

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij} x_{2ij} + \beta_4 x_{3ij} + \beta_5 x_{3ij}^2 + \beta_6 x_{3ij}^3 + \beta_7 x_{1ij} x_{3ij} + u_j + e_{ij} \quad (5)$$

$$u_j \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

where the response is the Year 5 score and the Year 3 score (x_{3ij}) is added as predictor as well as quadratic and cubic terms, and a further interaction between Indigenous status and

Year 3 score. The school is the Year 5 school.[9] The longitudinal model results (Table 6) for Reading and Numeracy use both methodological procedures as discussed, that is, imputation for missing Year 3 scores and adjustment for reliability, as does our final model.

Insert Table 6 about here

The first outcome from the results is at least a cubic relationship between Year 5 score and Year 3 score for both Reading and Numeracy for all students, indicating that progress across Year levels is not a simple linear change. In other words, as shown in Figure 1, the rate of change of Year 5 by Year 3 score is somewhat less for very high and very low Year 3 scores. The longitudinal analyses highlight different Indigenous and gender interactions from the cross-sectional analyses. In the cross-sectional analyses, Year 5 Numeracy outcomes had a significant gender-Indigenous status interaction. In the longitudinal analyses, only a small and non-significant interaction is apparent for Numeracy progress. However, for Reading progress, the difference between female and male Indigenous students is not as large as the gender difference for non-Indigenous students, although girls overall improve at a slower rate than boys in both Reading and Numeracy.[10]

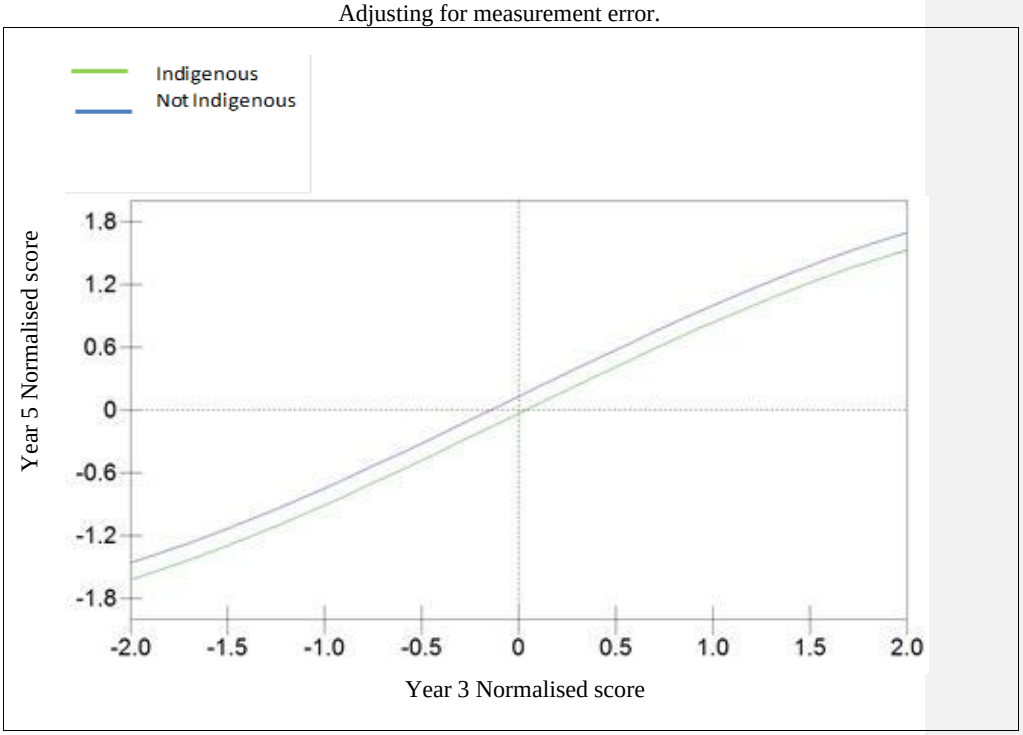
The analyses show that Indigenous students, after adjusting for Year 3 score, are predicted to do worse at Year 5 and in this sense to fall further behind. There is no discernible interaction whereby the relationship between Year 5 and Year 3 scores differs for Indigenous students. Figure 1 illustrates the predicted Year 5 Numeracy score for non-Indigenous boys,

[9] As noted, some 20% of students change school between Year 3 and Year 5 but this does not appear to be associated with progress between these ages. Fitting a cross-classified model to study the relative effect of the Year 3 school and Year 5 school on progress is of interest and a topic for further research.

[10] Both for Reading and Numeracy, the estimated Level 2 residuals u_j, the 'school effects', are little changed after adjustment for measurement error. The adjusted and unadjusted estimates are correlated 0.99 with similar standard errors. We have also fitted models where the coefficients of the Year 3 score and Indigenous status vary across schools. There is some evidence that the (random) coefficients do indeed vary across schools. However, the corresponding model fixed effects do not change markedly when such a model is fitted. Thus, while when comparing schools it would be important to take such effects into account, for present purposes we do not present this analysis.

with results displayed for the range of Year 3 scores between approximately the 2.5th and 97.5th percentiles, corresponding to the range (-2, 2) on the normalized Year 3 score scale.

Figure 1. Boys. Year 5 normalised Numeracy NAPLAN score against Year 3 normalised score. Adjusting for measurement error.



A similar pattern occurs for Reading for Indigenous and non-Indigenous boys.

Our final model (Table 7) examining student Reading and Numeracy progress incorporates all demographic variables used in national reports including school geographic location as a second-level variable and the language used in the home (but excluding parental education and occupation). As noted, location effects may be confounded by other factors including an interaction for Indigenous students in remote regions with language spoken in the home. Therefore, location becomes a proxy variable that may provide evidence of the impact of a number of factors on student achievement but further interpretation is beyond the analyses in this paper.

Table 7 shows the results of fitting the two level longitudinal model including geographic regions and interactions with Indigenous status to examine progress from Year 3 to Year 5, incorporating both our imputation process as well as adjustments for measurement error in the tests. As occurred with the previous model, in this model the relative amount of variation attributable to schools, as measured by the variance partition coefficient, is twice as high for Numeracy as for Reading, a finding that has been seen elsewhere (see, e.g., Goldstein et al. 1993).

Firstly, once more, a cubic relationship is found between Year 5 achievement and Year 3 score, indicating that progress across Year levels for all students is not a simple linear change. In terms of gender differences, girls overall make less progress than boys in both Reading and Numeracy. However, the difference in progress between female and male Indigenous students in Reading is still not as large as the gender difference for non-Indigenous students, while the interaction for Numeracy is negligible. The variable 'Language background other than English' shows that while students who have a language other than English spoken in the home make less progress in Reading than students with only English spoken in the home, the reverse effect on progress occurs, and is more marked for Numeracy. The effect of school location for all students is mixed. All students in provincial and remote/very remote areas make less progress in Numeracy than those in metropolitan areas, but not in Reading, perhaps indicating that an important area to examine should be numeracy teaching in non-metropolitan schools. While there is negligible interaction between Indigenous students and Year 3 score, there is a non-negligible interaction between gender and Year 3 score. Girls do not progress at the same rate as boys. An important interaction effect that emerges is between Indigenous status and school location, with the Indigenous students in remote/very remote locations making substantially less progress in both Reading

and Numeracy than Indigenous students in metropolitan or provincial locations. That is, Indigenous students in these regions who are behind in Year 3 fall further behind in Year 5.

Insert Table 7 about here

Discussion

Overall, these analyses provide more detailed insight into NAPLAN achievement than current national reports and analyses drawing on the available measures related to equity education goals used in national reports of schooling. The analyses confirm that, on average, Year 5 achievement in Reading and Numeracy on NAPLAN tests is highly influenced by prior Year 3 achievement, but not in a linear fashion, at least not on the chosen scale. Progress is lower for students who are lowest and highest achieving than for other students. The final model shows that girls make less progress than boys in both areas of achievement.

Having a language background other than English (language spoken in the home) appears to be associated with somewhat greater progress for Numeracy but less progress for Reading, with both developing from a lower achievement base, than for students who do not have another language spoken in the home. These core findings from our analysis and statistical approach provide greater insight than those presented in NAPLAN national reports both for cross-sectional and longitudinal (change based on average scores) reports. They demonstrate the value of undertaking more detailed analyses of a national database resource that entails a considerable financial and human resource commitment by education systems, schools, teachers and students. At present, the extent to which NAPLAN outcomes have been used to inform policy is not only limited in scope but underwhelming in education direction. They indicate areas where further research is needed to examine including why progress for girls slows and why students with a language background other than English in general are improving in Numeracy but not in Reading. The analyses also demonstrate that the

achievement of the students in non-metropolitan schools declined relative to their metropolitan peers from Year 3 to Year 5. Disadvantage is increasing.

Importantly, the analyses provide clearer information based on NAPLAN data on the achievement gap for Indigenous and non-Indigenous students. As regularly noted, in terms of achievement in Years 3 and 5, on average Indigenous students perform considerably less well than non-Indigenous students. The effect for progress is more marked for Indigenous boys than Indigenous girls for Reading and more marked for those in remote and very remote areas. A major insight our analysis provides is that not only does the 'gap' remain between the achievement of Indigenous and non-Indigenous Australian students, but that it increases over school years. Indigenous students from a remote or very remote background show markedly less progress than those from provincial or metropolitan areas, even more markedly in Numeracy. The available data do not allow further exploration of why this is occurring. It is not possible to examine interaction effects for Indigeneity, remoteness of location, and home language background due to the small number of students. However, the differential overall effects of LBOTE on progress are worth more exploration for Indigenous students with a language background other than English in the home. Given these findings, the effect is not transparent.

These analyses have both methodological and substantive implications. Methodologically, the present study demonstrates that use of advanced statistical models in combination with procedures to adjust both for missing values in longitudinal analyses and for measurement errors in the test scores do extend inferences beyond those provided in the official NAPLAN reports (e.g., ACARA, 2017). In addition to multilevel longitudinal analyses of individual student achievement and progress, approaches used in the study incorporate enhanced imputation procedures to address critical missing data, and adjustments for measurement error in NAPLAN tests. We have adopted what appear to be reasonable

values for measurement error variances, but the effect on the parameter estimates will depend on the values used. The methodological enhancements provide greater strength to interpretation of NAPLAN achievement data for all students and especially for Indigenous students. These enhancements have relevance beyond Australia and reporting of NAPLAN to similar accountability analyses in other countries, and international achievement test comparisons.

Implications for Policy and Practice

The substantive findings indicate that more research is needed to understand achievement and progress for all students in identified equity groups, including students who are evidenced to be disadvantaged by geo-location of their education. Further, despite policy and resource commitment, the gap between Indigenous and non-Indigenous student achievement has previously been identified as not decreasing. However, as our analyses show, the gap actually increases as Indigenous students move through schooling. It is an area still in need of further investigation and research to identify effective policy directions to improve Indigenous students' achievements and future life and work opportunities.

Importantly, however, not all Indigenous students have low achievement—Figure 1 and our analyses confirm that there are high-achieving Indigenous students who should progress at a similar rate to non-Indigenous students. It is important therefore that any policies introduced to improve Indigenous student learning should address learning for Indigenous students at all achievement levels. A major policy initiative to address the Indigenous student gap has been to focus on 'underachieving schools' as entities. As our analyses show, supporting earlier research findings, individual student prior achievement is the most important predictor of future achievement. Policies need to focus on "supporting those students who are under-performing or falling behind, not on a small proportion of schools with particular characteristics" (Marks 2015, p. 12). As Helme and Lamb (2011, p. 2)

noted, following a review of research, “a ‘one size fits all’ approach that either treats Indigenous students the same as non-Indigenous students or assumes that all Indigenous young people are the same” will not be successful.

Conclusion

The analytic approaches outlined in this article are applicable to longitudinal accountability datasets internationally. As the Australian analyses show, analytic methods are available for use with such data to provide greater insights into student achievement and progress to inform policy. With respect to the educational progress of Australian Indigenous students, the educational challenges faced by Australian Indigenous students, especially those in remote communities, are echoed internationally for Indigenous and other students with potential disadvantage in many countries. As in Australia, tendency to focus on *average* achievement of cohorts, rather than examining individual student learning patterns, can lead to policy and implementation of strategies that do not provide an equitable perspective on student learning. For example, as a starting point, based on the analytic outcomes presented in this paper, we recommend that future Australian policy should focus on low-achieving students and relative progress, whether for girls or boys, for Indigenous students, or students living in remote and very remote locations. However, we note that strategies to address the ‘gap’ also need to focus on high-achieving students from potentially disadvantaged backgrounds to ensure that they, as for other students, continue to be challenged.

Limitations of the Study

The analyses presented in this paper use the variables provided in a dataset by ACARA. Although the overall sample size for the Queensland data analyzed in this paper is large, datasets involving Indigenous students in some contexts are relatively small. Language background other than English has been used as a variable in the study, firstly, as it provides evidence of the impact of a changed interpretation for analyses when data are imputed rather

than excluded, and secondly, as it provides interesting outcomes in terms of Reading and Numeracy progress from Year 3 to Year 5. As we note above, an area worth further examination is the impact of language background other than English on the achievement of Indigenous students, especially those in schools in remote and rural geo-locations. However, given the small numbers of Indigenous students with other language backgrounds in these areas, and the concentration of these students in remote and rural geo-locations, as identified earlier, research into this issue needs to be undertaken through different approaches including qualitative and case study analyses.

A further limitation in the study is the lack of information on other contextual variables that may affect Indigenous student literacy and numeracy achievement on NAPLAN including specific evidence of cultural bias and effect of language, and the socio-educational advantage or disadvantage of Indigenous students, in terms of parental occupation and education, in different school regions. Finally, the data reflect only output measures of literacy and numeracy achievement measured through a standardized test. They do not enable exploration of the impact of different pedagogical interventions that may be occurring within schools to address learning improvement for students with limited literacy and numeracy, regardless of cultural heritage, language background or gender.

References

ACIL Allen Consulting with La Trobe University & Phillips KPA. (2014). *Evaluation of the Aboriginal and Torres Strait Islander Education Action Plan 2010*-2014: Final Evaluation Report. Brisbane, Qld: ACIL Allen Consulting.

Australian Bureau of Statistics. (2013a). *3101.0 – Australian demographic statistics, Jun 2013*. http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/61D19DFBD033B48FCA257CA7000DCB16?opendocument. Accessed 4 February 2019.

Australian Bureau of Statistics. (2013b). *Estimates of Aboriginal and Torres Strait Islander Australians, Jun 2011*. http://www.abs.gov.au/Ausstats/abs@.nsf/0/A03584675FA93CECCA25762A001D0701?OpenDocument. Accessed 4 February 2019.

Australian Bureau of Statistics. (2014). *4221.0-Schools, Australia, 2013.* http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4221.0main+features100062013. Accessed 4 February 2019.

Australian Curriculum, Assessment and Reporting Authority. (2013). *National Assessment Program—Literacy and Numeracy. Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2013*. Sydney, NSW: ACARA.

Australian Curriculum, Assessment and Reporting Authority. (2014). *National Assessment Program—Literacy and Numeracy. Technical Report. 2013.* Sydney, NSW: ACARA.

Australian Curriculum, Assessment and Reporting Authority. (2016a). *Assessment scales – NAPLAN*. https://www.nap.edu.au/results-and-reports/how-to-interpret/scales. Accessed 4 February 2019.

Australian Curriculum, Assessment and Reporting Authority. (2016b). *NAPLAN 2016 technical report.* Sydney, NSW: ACARA.

Australian Curriculum, Assessment and Reporting Authority. (2016c). *Why NAP*. https://www.nap.edu.au/about/why-nap. Accessed 26 September 2019.

Australian Curriculum, Assessment and Reporting Authority. (2017). *National Assessment Program—Literacy and Numeracy. Achievement in reading, persuasive writing, language conventions and numeracy: National report for 2017.* Sydney, NSW: ACARA.

Australian Curriculum, Assessment and Reporting Authority. (2018). *Test development*. https://www.nap.edu.au/about/test-development. Accessed 2 September 2019.

Australian Government. (2014). *Australian Government response to the Senate Education and Employment References Committee Report Effectiveness of the National Assessment Program – Literacy and Numeracy*. https://docs.education.gov.au/node/36183. Accessed 4 February 2019.

Australian Government, Australian Law Reform Commission. (2010). Legal definitions of Aboriginality. https://www.alrc.gov.au/publication/essentially-yours-the-protection-of-human-genetic-information-in-australia-alrc-report-96/36-kinship-and-identity/legal-definitions-of-aboriginality/ Accessed 9 September 2019.

Commonwealth of Australia, Department of the Prime Minister and Cabinet. (2018). *Closing the gap: Prime Minister's report 2018*. Canberra, ACT: DPMC.

Council of Australian Governments. (2009). *National integrated strategy for closing the gap in Indigenous disadvantage.* http://webarchive.nla.gov.au/gov/20130429035428/http://www.coag.gov.au/sites/default/files/NIS_closing_the_gap.pdf. Accessed 4 February 2019.

Council of Australian Governments. (2013). *National education reform agreement.* http://www.federalfinancialrelations.gov.au/content/npa/national_agreements/past/national-education-agreement.pdf. Accessed 4 February 2019.

Cumming, J., Jackson, C., Day, C., Maxwell, G., Adie, L., Lingard, B., Haynes, M., & Heck, E. (2018). *2018 Queensland NAPLAN review. School and system perceptions report and literature review. Bri*sbane, Qld: Institute for Learning Sciences and Teacher Education. https://qed.qld.gov.au/programsinitiatives/education/Documents/naplan-2018-school-perceptions-report.pdf. Accessed 2 September 2019.

Cumming, J., & Goldstein, H. (2016). From school to employment: Analysing the careers of Year 9 students using the Longitudinal Surveys of Australian Youth. *Longitudinal and Life Course Studies Journal*, *7*, 53–63. http://dx.doi.org/10.14301/llcs.v7i1.342

Goldstein, H. (2011). *Multilevel statistical models* (Fourth ed.). Chichester, UK: Wiley.

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), *177*(2), 553–564. https://doi.org/10.111/rssa.12022

Goldstein, H., & French, R. (2015). Differential educational progress and measurement error, *Journal of Longitudinal and Lifecourse and Studies*, *6*, 351–358.

Goldstein, H., Browne, WJ., & Charlton, C. (2018). A Bayesian model for measurement and misclassification errors alongside missing data, with an application to higher education participation in Australia. *Journal of Applied Statistics*, *45*(5), 918–931. https://doi.org/10.1080/02664763.2017.1322558

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society, Series* A, *170*(4), 941–954. https://doi.org/10.1111/j.1467-985X.2007.00491.x

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A multilevel analysis of school examination results, *Oxford Review of Education*, *19*(4), 425–433.

Goldstein H., & Wood R. (1989). Five decades of Item Response Modelling. *British Journal of Mathematical and Statistical Psychology*, *42*, 139–167.

Guenther, J., Bar, M., & Osborne, S. (2013). Red dirt thinking on educational disadvantage. *The Australian Journal of Indigenous Education*, *42*, 100–110. https://doi.org/10.1017/jie.2013.18

Helme, S., & Lamb, S. (2011). *Closing the school completion gap for Indigenous students*. Resource Sheet No. 6 for the Closing the Gap Clearinghouse, Australian Institute of Health and Welfare https://www.aihw.gov.au/getmedia/83074cc6-7dc9-40fb-9505-7b49c07e68e3/ctgc-rs06.pdf.aspx?inline=true. Accessed 4 February 2019.

Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education*, *20*(1), 77–93. https://doi.org/10.1080/10476210802681741

Klenowski, V., & Gertz, T. (2009). Culture-fair assessment: Addressing equity issues in the context of Primary mathematics teaching and learning. https://research.acer.edu.au/research_conference/RC2009/17august/15/. Accessed 2 September 2019.

Klenowski, V., Tobias, S., Funnell, B., Vance, F., & Kaesehagen, C. (2010). *Culture-fair assessment: Challenging Indigenous students through effortful mathematics teaching*, Paper presented at the AARE Annual Conference, Melbourne, 28 November to 2 December 2010.

Lingard, B., Thompson, G., & Sellar, S. (Eds.). (2016). *National testing in schools. An Australian assessment*. Oxon, UK: Routledge.

Marks, G. (2015). Students, schools and socioeconomic background. *Independence*, *40*(1), 4, 6, 8, 10, 12.

Ministerial Council for Education, Early Childhood Development and Youth Affairs. (2009). *Aboriginal and Torres Strait Islander education action plan* 2010 – 2014.

http://scseec.edu.au/archive/Aboriginal-and-Torres-Strait-Islander-Education-Action-Plan.aspx. Accessed 4 February 2019.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2006). *Australian directions in Indigenous education 2005-2008.* Carlton South, Vic: MCEETYA).

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008). *Melbourne declaration on educational goals for young Australians.* http://www.curriculum.edu.au/verve/_resources/national_declaration_on_the_educational_goals_for_young_australians.pdf. Accessed 4 February 2019.

Rice, S., Dulfer, N., Polesol, J., & O'Hanlon, C. (2016). NAPLAN and student wellbeing: Teacher perceptions of the impact of NAPLAN on students. In R. Lingard, G. Thompson & S. Sellar (Eds.), *National testing and its effects: Evidence from Australia* (pp. 72–85). Abingdon, UK: Routledge.

Senate Standing Committee on Education and Employment [SSCEE]. (2014). *Effectiveness of the National Assessment Program–Literacy and Numeracy. Final report*. Canberra, ACT: SSCEE.

Steering Committee for the Review of Government Service Provision. (2014). *Overcoming Indigenous Disadvantage: Key Indicators 2014*. Canberra, ACT: Productivity Commission.

Thompson, G., Adie, L., & Klenowski, V. (2017). Validity and participation: Implications for school comparison of Australia's national assessment program. *Journal of Education Policy*, 33(6), 759–777. https://doi.org/10.1080/02680939.2017.1373407

Thomson, S., De Bortoli, L., & Underwood, C. (2017). *PISA 2015: Reporting Australia's results.* Camberwell, Vic: ACER.

Thomson, S., Hillman, K., Schmid, M., Rodrigues, S., & Fullarton, J. (2017). *Reporting Australia's results PIRLS 2016*. Camberwell, Vic: ACER.

Thomson, S., Wernert, N., O'Grady, E., & Sima Rodrigues, S. (2017). *TIMSS 2015: reporting Australia's results*. Camberwell, Vic: ACER.

Warren, E., Young, J., & de Vries, E. (2007). Australian Indigenous students: The role of oral language and representations in the negotiation of mathematical understanding. In J. Watson & K. Beswick (Eds.), *Proceedings of the 30th annual conference of the Mathematics Educational Research Group of Australia* (pp. 775-784). Australia: MERGA Inc.. https://merga.net.au/Public/Publications/Annual_Conference_Proceedings/2007_MERGA_CP.aspx. Accessed 4 February 2019.

Table 1. Descriptive statistics for Queensland student population ACARA database (Year 5 records 2013; excluding missing responses; N = 53,038)

		Number	Percentage
Gender	Female	26236	49.5
	Male	26802	50.5
Ethnicity	Indigenous students	3748	7.1
	non-Indigenous	49290	92.9
Language Background	Language background other than English	5006	9.4
	Language Background English	48032	90.6
Region	Metropolitan	36764	69.3
	Provincial	14540	27.4
	Remote/very remote	1734	3.3
Gender x Ethnicity [%age of ethnic group]	Female Indigenous students	1865	49.8
	Male Indigenous students	1883	50.2
	Female non-Indigenous students	24371	49.4
	Male non-Indigenous students	24919	50.6
Ethnicity x Region [%age of ethnic group]	Metropolitan Indigenous students	1807	48.2
	Provincial Indigenous students	1294	34.5
	Remote/very remote Indigenous students	647	17.3
	Metropolitan non-Indigenous students	34957	70.9

Provincial non-Indigenous students	13246	26.9
Remote/very remote non-Indigenous students	1087	2.2

Table 2. Mean scaled NAPLAN scores Year 3 (N=47,506 Reading; 47,292 Numeracy) and Year 5 (N = 53,038 Reading, Numeracy) using ACARA QLD database (Scale scores transformed to common national scale with mean of 500 and range 0 to 1000 for 10 bands from Year 3 to Year 9)

		Year 3 (missing excluded)		Year 5	
		Mean	sd	Mean	sd
Reading	Total	404.44	88.91	498.10	68.62
Gender	Females	414.44	87.89	503.72	67.04
	Males	394.55	88.82	492.59	69.70
Ethnicity	Indigenous	342.38	77.75	447.60	62.05
	Non-Indigenous	409.03	87.96	501.93	67.57
Language Background	Other than English	397.13	89.55	489.33	71.57
	English	405.10	88.83	499.01	68.24
Region	Metropolitan	411.55	88.63	503.08	68.32
	Provincial	392.28	86.99	490.25	66.91
	Remote/very remote	356.61	85.23	458.12	70.01
Numeracy	Total	388.20	71.67	482.69	72.65
Gender	Females	384.57	67.12	477.01	67.97
	Males	391.78	75.72	488.25	76.54
Ethnicity	Indigenous	339.01	60.58	425.85	64.55
	non-Indigenous	391.79	71.09	487.01	71.40
Language Background	Other than English	386.21	79.10	484.30	88.51
	English	388.38	70.96	482.52	70.79
Region	Metropolitan	393.44	72.24	488.46	73.07
	Provincial	378.90	68.65	473.30	68.97
	Remote/very remote	355.50	67.95	438.97	70.64

Table 3. Mean Year 5 normalised NAPLAN Reading scores by whether missing Year 3 test score, language background, Indigenous status, gender. Standard errors in brackets.

	Non-Indigenous	Indigenous
Not missing Year 3 score	0.003 (.0046) N=44233	-0.782 (.0162) N=3273
Missing Year 3 score	-0.213 (.0143) N=5057	-1.089 (.0414) N=475
	Female	Male
Not missing Year 3 score	0.028 (.0062) N=23628	-0.129 (.0065) N=23878
Missing Year 3 score	-0.183 (.0195) N=2608	-0.383 (.0196) N=2924
	Language background English	Language background other than English
Not missing Year 3 score	0.044 (.0047) N=43569	-0.128 (.0167) N=3937
Missing Year 3 score	-0.239 (.0156) N=4463	-0.495 (.0297) N=1069

Table 4. Year 5 normalised NAPLAN Reading scores related to Year 3 scores, gender, Indigenous status and language background other than English. Standard errors in brackets. Imputation by MCMC burnin=150, Iterations=500. Well mixing chains

Parameter	Listwise deletion	Imputation
Intercept (β_0)	0.081	0.066
Year 3 score (β_4)	0.747 (0.003)	0.745 (0.003)
Indigenous (β_1)	-0.208 (0.017)	**-0.239** (0.015)
Female (β_2)	-0.008 (0.006)	0.002 (0.006)
Language background other than English (β_5)	-0.001 (0.011)	**-0.036** (0.011)
Indigenous x Female (β_3)	0.066 0.023)	0.077 (0.020)
Level 2 variance (σ_u^2)	0.019 (0.001)	0.021 (0.002)
Level 1 variance (σ_e^2)	0.379 (0.002)	0.381 (0.003)
Number of records used	47,506	53,058

Table 5. Year 3 and Year 5 Reading and Numeracy scores separately related to gender and Indigenous status. Standard errors in brackets. [N= ; no of schools=]				
Parameter	**Year 3**		**Year 5**	
	Reading	**Numeracy**	**Reading**	**Numeracy**
Intercept (β_0)	**-0.102**	**0.079**	**-0.074**	**0.106**
Indigenous (β_1)	**-0.540 (0.028)**	**-0.666 (0.029)**	**-0.610 (0.010)**	**-0.747 (0.028)**
Female (β_2)	**0.155 (0.009)**	**-0.143 (0.009)**	**0.140 (0.010)**	**-0.201 (0.009)**
Indigenous x Female (β_3)	**-0.025 (0.038)**	**0.060 (0.039)**	**-0.005 (0.039)**	**0.106 (0.039)**
Level 2 variance (σ_u^2)	**0.139 (0.006)**	**0.157 (0.006)**	**0.121 (-.006)**	**0.149 (0.006)**
Level 1 variance (σ_e^2)	**0.829 (0.006)**	**0.802 (0.006)**	**0.841 (0.006)**	**0.802 (0.006)**
Deviance (-2 loglikelihood)	**117035.9**	**116925.9**	**117559.5**	**117030.4**

Commented [A1]: Add numbers

Table 6. Year 5 normalised NAPLAN Reading and Numeracy scores related to Year 3 normalised Reading and Numeracy scores, gender and Indigenous status. Standard errors in brackets. N (Year 5) = 53,038, number of schools =1,369

Parameter	Reading	Numeracy
Intercept (β_0)	0.087 (0.007)	0.127 (0.008)
Year 3 score (β_4)	**1.005** (0.006)	0.**900** (0.006)
Year 3 score squared (β_5)	0.**007** (0.003)	-0.003 (0.003)
Year 3 score cubed (β_6)	-0.**053** (0.002)	-0.**028** (0.002)
Indigenous (β_1)	-0.**156** (0.021)	-0.**162** (0.019)
Female (β_2)	-0.**023** (0.006)	-0.**076** (0.006)
Indigenous x Female (β_3)	0.**087** (0.021)	0.027 (0.021)
Indigenous x Year 3 score (β_7)	0.003 (0.014)	0.007 (0.013)
Level 2 variance (σ_u^2)	0.022 (0.001)	0.047 (0.003)
Level 1 variance (σ_e^2)	0.286 (0.002)	0.250 (0.002)

Table 7. Year 5 normalised Reading and Numeracy scores related to Year 3 normalised scores, gender, Indigenous status, geographic region of school and home language background. Standard errors in brackets. Adjusting for measurement error in Year 3 and Year 5 scores with reliability = 0.85. MCMC burn-in =250, iterations = 1000. N (Year 5) = 53,038 number of schools =1,369

Parameter	Reading	Numeracy
Intercept (β_0)	0.103 (0.009)	0.153 (0.010)
Year 3 score (β_4)	**1.033** (0.007)	0.**918** (0.006)
Year 3 score squared (β_5)	0.**009** (0.003)	-0.**007** (0.003)
Year 3 score cubed (β_6)	-0.**056** (0.002)	-0.**029** (0.002)
Indigenous (β_1)	-0.**112** (0.024)	-0.**132** (0.022)
Female (β_2)	-0.**034** (0.007)	-0.**083** (0.007)
Language Background other than English	-0.**027** (0.011)	0.**068** (0.011)
Provincial city	-0.010 (0.013)	-0.**032** (0.016)
Remote/very remote	-0.051 (0.030)	-0.**101** (0.036)
Indigenous x Female (β_3)	0.0**50** (0.023)	0.002 (0.022)
Indigenous x Year 3 score (β_7)	-0.003 (0.015)	-0.010 (0.013)
Female x Year 3 score (β_8)	-0.**042** (0.007)	-0.**033** (0.006)
Indigenous x Provincial	-0.034 (0.025)	-0.012 (0.024)
Indigenous x Remote/very remote	-0.**221** (0.043)	-0.**306** (0.042)
Female x Provincial	0.016 (0.013)	0.004 (0.012)
Female x Remote/very remote	0.027 (0.033)	0.014 (0.031)
Level 2 variance	0.021 (0.001)	0.043 (0.002)
Level 1 variance	0.286 (0.002)	0.250 (0.002)