Standardised Test Scores and Educational Achievement in Australia

Sarah Brittany Lehmann Cornell-Farrow

A thesis submitted in fulfilment of the

requirements for the degree of

Doctor of Philosophy

in the
School of Economics
Faculty of the Professions
The University of Adelaide

15 March 2019

Standardised Test Scores and Educational Achievement in Australia

Sarah B. L. Cornell-Farrow

15 March 2019

THE UNIVERSITY of ADELAIDE

Table of Contents

ABSTRACT

This thesis is comprised of four self-contained papers utilising standardised test score data, specifically data from the Australian National Assessment Program - Literacy and Numeracy (NAPLAN). The analysis presented by this thesis is particularly pertinent given the recent publication of the findings of the Review to Achieve Educational Excellence in Australian Schools, chaired by David Gonski AC.

The first paper provides a survey on the uses of standardised test score data for economic analysis, and a discussion of the uses and limitations of the NAPLAN data set. This paper establishes two main areas for analysis. Firstly, the analysis of school funding policy, given the provision of schools' financial data enabled by the My School website, which will be explored in the second paper of this thesis. Next, the analysis of how student background characteristics may impact achievement, which is analysed further in the third and fourth papers of this thesis.

As outlined above, the second paper focuses on the topic of school funding. This paper explores the causal impact of school funding on student achievement in NAPLAN. Using school-average test score data paired with funding information for each school, we determine how the three different types of funding received by Australian schools impact test scores differently depending on sector and state. We find, in general, that funding from the federal government has the least beneficial impact, with state government funding and parent fees more likely to provide the greatest benefit to schools. These results have a significant policy impact, indicating that funding is most beneficial when provided at as local a level as possible.

The third paper of this thesis turns to the socio-educational determinants of educational achievement in Australian schoolchildren. We find that students with an Indigenous or language other than English background are at risk of poor performance, as well as students with a parent who did not complete year 12, does not have a university degree or is not employed. Secondly, we find that private schooling makes a student more likely to meet and surpass national benchmarks for achievement, on average. However, the probability of a private school student performing in the higher NAPLAN bands changes based on their other socio-educational features.

This thesis concludes with a short fourth paper that provides another perspective to predicting the event of 'low achievement' by implementing machine learning strategies.

Together, these papers constitute an overview of the possibilities for econometric analysis of the NAPLAN data.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Sarah B. L. Cornell-Farrow
15 March 2019

ACKNOWLEDGMENTS

This thesis is dedicated TO MY GRANDMOTHER, Janet Farrow.

Thank you to Firmin Doko Tchatoka for the guidance, friendship, and support you have provided me with throughout this process as my principal supervisor. I will be forever grateful to you for giving me the confidence to do further research and to challenge myself intellectually. Thank you also to Virginie Masson for the advice and support you have given me during my time at the University of Adelaide. Special thanks to Prof. Gareth Myles and Stephanie McWhinnie for going above and beyond to support my studies. Robert Garrard, thank you for working with me on the final paper of this thesis and for always being willing to discuss an (and the) economic problem with me. I am further indebted to Timothy Hersey, Qazi Haque, Sean Parry, Kim Wu, and Mia Tam for help editing my manuscript, troubleshooting technology, navigating life as a PhD student, and for being all round great colleagues.

On a practical note, thanks must be given to the Australian Curriculum, Assessment and Reporting Authority for the provision of the data which made this thesis possible. This data set is confidential but may be obtained by researchers through ACARA's Data Access Program at http://www.acara.edu.au/contact-us/acara-data-access.

I am also grateful for the support I have received from an Australian Government Research Training Program (RTP) Scholarship, and the University of Adelaide Baillieu Supplementary Research Scholarship, which allowed me to dedicate my time to writing this thesis. Many thanks must also go to those who have provided me feedback at the University of Adelaide School of Economics and the CSIRO, at the 2016 Australasia Meeting of the Econometric Society, the 2018 Australian Conference of Economists, and the 31st PhD Conference in Economics and Business, as well as to a number of anonymous referees.

Special mention must be made of the formative time I spent as an intern in the Office of the Hon. Julia Gillard, the Minister for Education and then Prime Minister whom championed the implementation of the National Assessment Program - Literacy and Numeracy. This period set the path that ultimately led to me embarking upon this exercise. All ideas expressed herein are, of course, my own or those of my co-authors and do not reflect those of Ms Gillard or her office.

Finally, all my love to my family and friends for the support that you have provided me, and to my four-legged writing buddies; Stella, Macy, Murphy, Bella, and Harry (all errors and omissions are theirs).

> There's a saying in Illinois I learned when I was down in a lot of rural communities. They said, 'Just weighing a pig doesn't fatten it.' You can weigh it all the time, but it's not making the hog fatter. So the point being, if all we're doing is testing and then teaching to the test, that doesn't assure that we're actually improving educational outcomes.
>
> *Barack Obama*
> *44th President of the United States of America*

Preface

The turn of the millennium brought about a boom in the administration of standardised tests by governments across the globe, and thus increased the collection of data on educational outcomes and associated student background characteristics. The focus on the role of data in accountability practice has led to an explosion in the availability of data to not only the policy-maker, but also the econometrician. New technologies have meant that data can be made available quickly, with greater coverage and scope, and potentially in new forms not conducive to traditional econometric methods. Numerous projects have developed for the collection of education data, such as the My School website in Australia, the Barnard Columbia No Child Left Behind Data Project, as well as projects in developing countries, which can be explored at Open Data for Africa. We are also seeing a growing interest in data linkage projects, where, for example, school education data could be linked to other administrative data sets such as health records, census data, and government expenditures. In the modern age, data is no longer a scarce resource. A key question is then how this data can be used practically to *actually* improve policy outcomes.

The goal of the applied economist is primarily to identify the causal effect of one variable on an outcome variable. This can be particularly difficult in the case of observational data sets where data comes from administrative or survey sources, rather than collected from a controlled experimental study. Frequently, the type of data available to the applied economist is from administrative sources, such as a government census, or health records. The issue with these types of data is that it is more difficult to provide convincing arguments for the exogeneity of the variables it contains. Furthermore, it is challenging to provide counterfactual analysis of potential outcomes given that we only observe each unit of observation in one state of the outcome of interest, to which they were not randomly assigned. The analysis of school education is then further complicated as the variables of interest; for example, ‘achievement’, are abstract in nature and thus difficult to measure. It becomes difficult to develop well-justified theories that explain the assumptions underlying education theory and thus any modelling decisions.

Analysis of school education typically relies on empirical work rather than theoretical work. Moreover, school education has always been a popular field for the testing of novel econometric methodologies and theories to achieve causal inference. No such clearer example exists than the classic work on instrumental variables of Angrist and Krueger (1991). Consider the basic linear regression model estimated using ordinary least squares:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{1}$$

where y is the log of weekly earnings, x is years of schooling, ε is an error term, and $\beta_j, j = 0, 1$, are coefficients to be estimated with β_1 being the returns to schooling. For consistency of β_1 we require, among other assumptions, *exogeneity* of the covariates x, or $\mathbb{E}[e_i|x_i] = 0$. This assumption is often implausible, especially in the case of schooling. For example, it is unlikely that funding is determined completely exogenously,[1] and it has been well-documented that the choice of private schooling is not exogenous.[2] The standard approach to consistently estimating β_1 in the case of endogenous regressors is to use an instrumental variable. That is, an additional variable, z_i, that is uncorrelated with the outcome variable (except in its ability to predict x_i), yet is highly correlated with the endogenous covariate in question.

The Angrist and Krueger (1991) approach is to use the individual's quarter of birth, QOB_i, to instrument for their level of schooling.[3] This choice is based on institutional factors, where schooling in the United States of America was compulsory until the age of 16. The total schooling received by each individual therefore depended on their birthday. Those born earlier in the year are older than their classmates when they begin school. This means that if they drop out as soon as they turn 16 they will have completed fewer years of schooling than their classmates born later in the year. These younger individuals have to wait longer until they turn 16 to drop out, thus accruing more time in schooling. Angrist and Krueger (1991) argue that *quarter of birth* is therefore correlated with *schooling* (*i.e.*, relevant). This variable satisfies the exclusion restriction as they argue that an individual's birthdate is unlikely to be correlated with other personal

[1]This is explored in the second paper of this thesis.

[2]We look at this difficulty in estimation in the third paper of this thesis.

[3]Other attempts to identify the returns to schooling include Card (1993), Lemieux and Card (2001), and approaches utilising data on twins such as Ashenfelter and Krueger (1994). See Card (1999) for a review of the education and earnings literature.

attributes and *wage* (*i.e.*, it provides *exogenous* variation in schooling).[4] Using two-stage least squares (2SLS), they find that men who attained more schooling as a result of these compulsory schooling laws received higher wages. Each extra year of schooling gives an estimated wage increase of about 7.5%. They particularly highlight the implication that their findings have for the literature on omitted variable bias; arguing that conventional OLS estimates of the return from schooling are biased downwards.

While this appears to be a simple yet effective approach, it is not without its limitations. Bound et al. (1995) discuss the issues inherent to instrumental variable estimation when instruments are *weak*, that is, when the correlation between the instrument(s) and the endogenous explanatory variable is low. In fact, they argue that this case is common. This presents a danger to applied work, as even in the presence of what the researcher sees as a clear 'natural experiment' or source of exogenous variation from a policy sense, it is not enough to justify that IV estimates are unbiased.

The discussion above provides just one example of the significant difficulties involved in any econometric analysis of education. The central question of this thesis is, given the multitude of test score data becoming available to the researcher, if and how can this data be used effectively in the econometric analysis of education policies. As highlighted by Varian (2014) and Einav and Levin (2014), it has been taken for granted that 'big data' sets, or any large modern data sets, will dramatically change how businesses and governments operate. Large-scale administrative data sets are purported to provide the opportunity to improve the ways we analyse any kind of economic activity. We believe that

[4] Note that Bound et al. (1995) invalidated the quarter of birth IVs, and many other studies have found them to be weak.

the collection of standardised test score data in Australia provides an interesting case study for these massive data sets, electronic versions of which are highly available to researchers. The goal of this thesis is to provide some examples of potential econometric applications of National Assessment Program - Literacy and Numeracy data to education policy in Australia.

The 'Gonski' reports on schooling and the rhetoric surrounding them have provided a set of assumptions for schooling policy in Australia that have underpinned a very particular approach to combatting the problem of students achieving poor educational standards. These include that funding can be a cure for poor performance, and a remedy for social disadvantage. It also reinforces the presumption that all students can perform 'above average', which we know to be a mathematical impossibility. In order to truly "ensure that differences in educational outcomes are not the result of differences in wealth, income, power or possessions," (Gonski et al., 2011) we must question these assumptions properly before providing blueprints for policy.

The contribution of this thesis is to explore these assumptions and provide a discussion of what analysis is appropriate, and in turn what conclusions can and can not be made, from the NAPLAN set of data. The papers here contained provide, to our knowledge, not only the first systematic study of the effects of school funding on NAPLAN scores, but also the first systematic study of the determinants of poor achievement in students across Australia in both public and private schools. The first paper of this thesis provides a survey of existing approaches to the economic analysis of standardised test score data, and describes the NAPLAN testing programme and data. While the paper outlines a number of limitations inherent in the data that restricts the ability of the econometrician to achieve causal inference, it pinpoints two particular areas for further analysis, which will comprise the main body of the thesis. First, to use the financial information in the school-level data to develop a better understanding of how funding relates to achievement in Australian schools; and second, to use the student-level data to examine what kinds of background characteristics make a student at risk of poor performance.

With this agenda in mind, the second paper questions the assumption that funding is causally related to test scores. We argue that the difficulty in estimating the causal effect of funding is twofold. First, funding may affect outcomes in a heterogenous manner over the distribution of test scores, and therefore can not be identified adequately with a mean-focused approach. Given this, we implement the quantile regression for panel data (QRPD) method (Powell, 2014, 2015, 2016) to estimate these effects. Second, funding may be endogenously determined by test scores in the previous period. In order to achieve identification of this parameter, we propose instrumenting with the level of funding in the previous period.

The third paper of this thesis focuses on another source of bias that must be considered when analysing school education data, that is, that school choice suffers from sample selection. As with the Angrist and Krueger (1991) example, unobservable variables such as *ability* are a significant determinant of test scores. Moreover, school choice may often be determined by the same unobservables. We illustrate an alternative approach to the typical instrumental variable 2SLS method to address this sample selection bias, namely the control function approach of Wooldridge (2015), in an attempt to identify the effect of private schooling. These papers highlight the difficulties involved with the economic analysis of test score data and propose potential solutions for these issues, given the data currently at hand.

The final paper of this thesis explores machine learning approaches to prediction, with the goal of determining if these methods can improve the econometrician's ability to analyse problems of this nature. While the econometric analysis of standardised test score data remains an area for contention, with much left unanswered, the problems involved with identifying a causal effect should not prevent us from analysing the data as best we can. Using machine learning techniques, we attempt to predict if a student will perform below standard on their next NAPLAN test. We find that we are able to categorise these children with a simple logistic regression, and that machine learning techniques are not necessary to improve the quality of predictions.

This thesis attempts to provide a way forward to make the data collected from NAPLAN truly valuable to the policy-maker, and therefore to the children required to sit these tests, in order that it can assist in *actually* improving educational outcomes for Australian children.

Statement of Authorship

On standardised test scores in Australia: What can we learn?
Unpublished and unsubmitted work written in manuscript style

Author

Name: Sarah Cornell-Farrow
Overall percentage (%): 100%
Certification: This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

25 January 2019

On standardised test scores in Australia: What can we learn?*

Sarah Cornell-Farrow†

The University of Adelaide, School of Economics, Adelaide, SA, Australia, 5005.

Abstract

Understanding effective uses for standardised test score data is central to the development of good policy and practice. This study provides a discussion of the Australian National Assessment Program - Literacy and Numeracy (NAPLAN). We begin with an overview of standardised testing, and various programs administered to students across the globe. Next, we outline the NAPLAN testing programme in detail, followed by a discussion of the limitations of the resulting data. We conclude with a survey of the current literature analysing standardised test score data. We highlight two areas for future research in the Australian context: the analysis of school funding, and the analysis of social disadvantage in schooling.

Keywords: Education, national testing, standardised test score data, economic methodology, Australia
JEL Classification: B41, C81, I21.

1 Introduction

Standardised testing plays a primary role in designing policies aimed at improving education systems across the globe, as achievement on these tests is strongly correlated with measures such as national GDP and individual economic well-being (Barro, 1991; Peterson and West, 2003). In Australia, the federal government introduced the National Assessment Program - Literacy and Numeracy (NAPLAN) in 2008. These tests produce scores in literacy and numeracy for all

*This research is supported by an Australian Government Research Training Program (RTP) Scholarship. Thanks must also be given to the Australian Curriculum, Assessment and Reporting Authority for the provision of the data discussed by this study.

†Email: sarah.cornell-farrow@adelaide.edu.au

Australian students in grades 3, 5, 7, and 9 each year, which are used regularly to inform education policy. Despite efforts to improve schooling, the performance of Australian students in international tests, such as the Programme for International Student Assessment (PISA), has been declining.[1] This paper therefore aims to investigate the uses and limitations of the NAPLAN test score data, in an effort to understand what we can learn in Australia about improving education standards using standardised tests.

The application of econometric techniques to test scores is a popular and growing field internationally. Modern analysis of the Australian education sector has taken the form of two reviews, undertaken by independent committees led by businessman David Gonski AC to provide advice to the federal government. The 2011 *Review of Funding for Schooling* (Gonski et al., 2011) and the 2018 *Review to Achieve Educational Excellence in Australian Schools* (Gonski et al., 2018) have outlined numerous deficiencies in Australian schooling and attempted to propose solutions for these. However, many of these solutions are based on beliefs about schooling policy that are in some cases disproven or contentious in the economic literature. For example, there is an emphasis on improving outcomes for disadvantaged students by increasing their funding, yet Hanushek (2003) has shown that this relationship may be intangible. Moreover, the most recent review has proposed that NAPLAN testing be made redundant by replacing it with an online assessment tool for teachers, arguing that NAPLAN is inflexible and timed poorly. It is therefore at present particularly important to understand the usefulness and effectiveness (or ineffectiveness) of NAPLAN testing to argue if it should or should not be replaced. This paper will recommend uses for the entire NAPLAN data set to understand how we can best

[1] See Ryan (2013) for a discussion of the Australia's declining PISA test scores.

utilise standardised test results to improve schooling policies in Australia.

This paper assesses the international literature on standardised test scores, in order to find what the Australian data can contribute. We find that, compared to the kinds of analyses that have been implemented internationally, there are significant limitations inherent to the NAPLAN data set. For example, IV approaches and panel data approaches are difficult to implement given the current state of the data. However, there remain two policy areas that the applied economist can learn about in Australia: (1) the funding of schools, and (2) the equitability of schools.

The paper will proceed as follows. Section 2 details standardised test score data sets from across the globe that have been used for economic analysis. Section 3 focuses on an overview of the Australian NAPLAN data set followed by a discussion of its limitations in Section 4. Section 5 provides a review of the existing literature. Section 6 concludes.

2 Standardised testing

Standardised testing is used by many countries as a means for evaluating their education systems.[2] A summary is given in Table 1. Perhaps the most well-known examples of testing programs are those undertaken internationally; namely, the Programme for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS), and Trends in International Mathematics and Science Study (TIMSS). These international tests allow comparisons between countries, enabling policy-makers at both national and international levels to design large-scale policy goals for education. For example, the inform-

[2] Morris (2011) provides an overview of standardised testing in OECD countries.

ation may be used to address Millennium Development Goals (MDGs) or Sustainable Development Goals (SDGs) by the United Nations. Moreover, they are used by countries to determine how their education system may be performing relatively to those across the globe. The primary role of these tests is therefore that of education policy evaluation.

Conversely, national testing programs may perform many different functions. Governments primarily introduce them for the purpose of monitoring school systems and policy design, at a more detailed national level. For example, the No Child Left Behind Act of 2001 (NCLB)[3] was instituted in the United States of America to enforce compulsory national testing (Peterson and West, 2003). The aim of NCLB was to require states to set achievement standards against which all students would be tested, in an effort to combat disadvantage in the educational system. The testing process was tied to federal grants, as well as certain policy interventions that were aimed at improving schools missing their targets. Testing programs such as these are known as 'high-stakes' testing, that is, testing that has significant consequences for either the student taking the test or the school.[4] Other examples of high-stakes testing regimes include the Pan-Canadian Assessment Program (PCAP), which has consequences for some high school students' final grades and university entrance, and the System to Measure Quality in Education (SIMCE) in Chile, which rates school quality based on their achievement.

Other types of testing programs, or 'low-stakes' programs, may still play important policy roles. For example, they still measure student and school achievement in a way that may monitor school performance. Importantly, they may also

[3]Replaced by the Every Student Succeeds Act in 2015.
[4]See McDermott (2011) for more on high-stakes testing.

Table 1: Summary of standardised testing programs

Data source	Country	Time period	Subjects	Age of participants	Sample
Programme for International Student Assessment (PISA)	OECD	2000, 2003, 2006, 2009, 2012, 2015, 2018	Science, reading, and math	15 year olds	Randomly selected schools in participating countries
Progress in International Reading Literacy Study (PIRLS)	International	2001, 2006, 2009, 2012, 2015, 2018	Reading	Grade 4	Randomly selected schools in participating countries
Trends in International Mathematics and Science Study (TIMSS)	International	1995, 1999, 2003, 2007, 2011, 2015, 2019	Math and science	Grades 4 and 8	Randomly selected schools in participating countries
National Assessment Program - Literacy and Numeracy (NAPLAN)	Australia	Annually since 2008	Reading, writing, grammer and language conventions, spelling, numeracy	Grades 3, 5, 7, and 9	Census
Pan-Canadian Assessment Program (PCAP)	Canada	2007, 2010, 2013, 2016	Mathematics, reading, writing, and science	Grade 8	Random sampling
System to Measure Quality in Education (SIMCE)	Chile	Since 2012 (but earlier test data available)	Many subject areas	Grades 2, 4, 6, 8, 10, and 11	Census
National Assessment of Educational Progress (NAEP)	USA	Various forms since 1969 (NCLB Act enforced 2001)	Many subject areas	Grades 3-8	Representative samples/census
National Monitoring Study of Student Achievement (NMSSA)	New Zealand	1995-2010 (NEMP) then NMSSA from 2012	Many subject areas	Grades 4 and 8	200 randomly selected schools
National Literacy and Numeracy Strategy	Ireland	Since 2012	Reading and mathematics	Grades 2, 4, and 6	Census
National Achievement Survey (NAS)	India	2001-2017	Mathematics, languages, science, and social sciences	Grades 3, 5, 8, and 10	Random sampling of students within selected schools
National Institute for the Evaluation of the Educational System of Education and Training (INVALSI)	Italy	Since 2009	Mathematics and Italian	Classes in primary and secondary school	Census
Southern and Eastern Africa Consortium for Monitoring Educational Quality	16 African countries	SACMEQ I 1995-1999, SACMEQ II 2000-2004, SACMEQ III 2006-2011	Reading and math	Grade 6	Various numbers of participating countries in each wave

Note: For further information about national testing in European countries see Eurydice, European Commission play a role in *formative assessment*, that is, they behave as diagnostic tests to allow teachers and schools to isolate students and subject areas that need increased focus or resources. Testing regimes such as the National Institute for the Evaluation of the Educational System of Education and Training (INVALSI) testing in Italy, the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) tests in Africa, the National Achievement Survey (NAS) in India, the National Literacy and Numeracy Strategy in Ireland, and the National Monitoring Study of Student Achievement (NMSSA) in New Zealand are examples of low-stakes testing regimes.

It must be noted that some researchers question whether test scores exhaustively measure educational achievement. Discussions of this nature lie outside the scope of this study and will be left to those from the field of education. Jacob et al. (2014) discuss the use of aggregate data in the evaluation of schools, and conclude that these kinds of data are in fact sufficient and appropriate to evaluate a range of educational policies. Furthermore, these kinds of data are typically accepted in the economics literature as a valid measure of educational output. Hanushek and Woessmann (2011) have also shown that these kinds of data are important for regressions involving economic growth, despite their potential flaws.

3 NAPLAN testing

The first round of NAPLAN testing took place in 2008, and since has been administered yearly to students in years 3, 5, 7, and 9 across all schools in Australia. The NAPLAN tests cover reading, writing, numeracy, and language conventions (spelling, grammar, and punctuation). The questions in all tests typically range in difficulty in order to best diagnose the level that each student is at. There are also particular skills that a student is expected to have mastered at certain year levels (*e.g.*, to recognise and continue a number pattern), and questions allowing them to demonstrate their ability to meet these minimum standards also form part of each test.

The reading test focuses on the reading and comprehension of English. Students are provided with a magazine containing various types of written text. They then answer questions related to the text. An example question could be (following reading a text about gardening): *Which word or group of words from the last paragraph tells the reader when to take the vegetables out of the garden?*[5] The results from this test are those typically analysed by researchers looking at 'reading skills'.

The numeracy test assesses a student's ability in mathematics, namely: number and algebra; measurement and geometry; and statistics and probability. The tests contain both multiple choice and short answer questions. In years 7 and 9, the students sit both a section where they are allowed a calculator, and another section where they are not. An example question from a numeracy test might be: *Select the two even numbers from the list below.* This is another test that is of particular interest to researchers, in order to understand mathematical achievement.

NAPLAN consists of two other tests, which are less popular for analysis by researchers. The first is a writing test that provides students with a prompt about which they must write a certain type of text (imaginative, informative, or persuasive). An example for an imaginative prompt may be: *The idea for your story is 'The Box'.* A persuasive prompt may be: *Reading books is better than*

[5]Example questions are taken from nap.edu.au

watching TV. There has not yet been a test assessing informative writing. Due to these differences in text type, it is more difficult to compare scores for the writing test across testing periods, even for the same student. The writing test also differs in that it must be assessed by graders against criteria (rather than there being correct answers). While the writing test measures an important skill and is of great value, it is of most use to schools and teachers, not to policy-makers, and will not be analysed further here.

The final NAPLAN test assesses language conventions, that is, spelling, grammar, and punctuation. This test complements the reading and writing tests where these skills must be demonstrated in context. There are separate minimum standards for spelling, and grammar and punctuation. The standards for grammar and punctuation might be; for example, identifying the correct location of a full stop. For spelling, a student may be able to demonstrate that they can correct spelling errors in a piece of text.

Pugh and Foster (2014) have provided some discussion on the relationship between the Australian NAPLAN data and data collected by other programs, but at that early point in time regarding NAPLAN testing there was little availability of data to researchers aside from that published publicly on the My School website, particularly regarding the individual student-level data. The NAPLAN data primarily differs from other data sets described in the previous section in that it cannot neatly be defined as 'low-stakes' or 'high-stakes' testing, which may provide difficulties in understanding the reliability of the data. While there are no clear consequences to poor performance on NAPLAN testing, there remains the threat of governments allocating funding based on scores, as well as the publication of a school's scores on the My School website, which may behave as a kind of 'hidden high stake' as it may affect school enrolments. The tests are described as benefiting students, parents, teachers, and schools, particularly in understanding a child's progress and identifying goals for teaching programs and uncovering those students that may need extra support. The tests are also of benefit to policy-makers, as the data can be used to support school improvement.

Raw data from NAPLAN testing, both student and school-level, for each year is available from the Australian Curriculum, Assessment and Reporting Authority (ACARA) upon application via their Data Access Program. School-average data is published publicly on the My School website each year.

4 Limitations of the NAPLAN data set

There remain a number of issues that limits the value of the NAPLAN data to the econometrician, particularly at the individual-level. Namely, (1) the lack of unique student identifiers and (2) background covariates that are both time-invariant and not detailed enough.

Firstly, the individual-level data lacks unique student identifiers that are consistent over time, making it difficult to follow a student between schools.[6] This makes it difficult to undertake certain kinds of analysis. For example, if a student changes schools between the previous testing period and the current period, the researcher has no information as to what school, or even school system, the student was previously in. It is therefore impossible to track switchers between the public and private school systems, which would provide better evidence as to the value (or lack thereof) of private schooling in Australia. Unique identifiers would also make it easier to link NAPLAN data to other administrative data

[6]Note that improving the ability to track students over time has been recommended as early as Miller and Voon (2011), and the introduction of a unique student identifier was a recommendation made by the most recent Gonski report.

such as health data or tax data to widen the types of analyses that are possible.

The second distinct limitation of the NAPLAN data is a lack of background covariates that would allow researchers to better account for socio-economic effects. This is despite the fact that governments have made it a priority that NAPLAN scores only be assessed in terms of schools with similar backgrounds. While there are ethical issues involved with the use of data on children, which cannot be denied, it would significantly improve the accuracy of any research using the NAPLAN data to provide more detailed covariates to the researcher. For example, the school-level data includes postcodes but this information is not included in the individual-level data. The data has been de-identified to the point that it almost lacks value completely, despite the requirements of ethics approval involved in the process of obtaining the data. There is scarce information as to the location of the school, which would provide a potentially valid and useful instrument for understanding the role of private schooling, as an example. The data set does not even differentiate between the Independent and Catholic schooling sectors, despite this differentiation at the school-level. These are just a few pieces of missing variables that limit the present capabilities of this data.

Both the Longitudinal Study of Australian Children (LSAC) and the Longitudinal Surveys of Australian Youth (LSAY) provide evidence of the usefulness of richer data sets including test score outcomes. For example, the LSAC provides data on NAPLAN scores alongside other non-cognitive scores, and more detailed family background information tracked over time. For example, Kalb and van Ours (2014) are able to estimate the effect of parents reading to children at age 4-5 on their test outcomes at least to age 10 or 11, which they find to be positive and significant. Furthermore, Warren and Haisken-DeNew (2013) use this data to show that pre-schooling (and the qualifications of their pre-school teachers) has a positive effect on a student's later educational attainment. This is all crucial information that is left out of the wider NAPLAN data set. As another example, the LSAY data is able to link PISA scores of a student at age 15 to their later employment rates and earning capacities at age 25 (Polidano and Ryan, 2017). Using data from the Victorian Government and TIMSS, Ryan (2017*b*) find that almost 10% of the variation in achievement in Australian high schools can be explained by the effect of teachers. Yet, both the school-level and individual-level NAPLAN data sets do not even discern which particular classroom a student is in within a school grade, let alone who they are being taught by. While the LSAC and LSAY remain valuable data sets, they both face their own issues of lower sample sizes, and attrition. The NAPLAN data collected by ACARA therefore provides a unique opportunity to broaden the evidence base for policy; however, the quantity of data collected must not be at the detriment of the quality of the variables included.

In their data survey on *PanelWhiz and the Australian Longitudinal Data Infrastructure in Economics*, Hahn and Haisken-DeNew (2013) describe the hierarchical structures of data sets such as the LSAC. They describe certain naming schemes for these hierarchical data sets as the 'international state of the art', and efforts to build a graphical user interface in STATA to easily analyse data sets that fit this structure. As the NAPLAN data is hierarchical in nature, *i.e.* state-level, school-level, and student-level, it begs the question as to why this data has not been formatted in such a way to gain the most use from it, as is the norm of other data sets. By presenting the data hierarchically, one could also include information on teachers,[7] and further variables that need not necessar-

[7] It is worth noting that the most recent report on schooling from the Grattan Institute also recommends collecting more data on teaching in order to support a more *adaptive* education system (Goss and Sonnemann, 2018).

ily fit the typical panel data structure. The data organisational tools exist, and modern machine learning techniques are suited to these kinds of data. Instead, the researcher analysing the NAPLAN data set is faced with a mammoth task in reorganising the structure of the data sheets available, with little payoff and ability to understand the hierarchical nature of the data, and thus the theoretical nesting of models.

While there are a number of limitations in the uses of the NAPLAN data set, it nevertheless provides a unique opportunity with which to research education policy in Australia. The potential uses of this data will be detailed in the following section.

5 Uses for the NAPLAN data

We identify five particular areas in which the economist may learn from Australia about schooling, using the NAPLAN data.[8] These areas cover test scores; (1) as accountability mechanisms; (2) as measurements of school funding efficiency; and (3) as metrics to analyse school management practices. These areas all fall under the umbrella of school resourcing. Next, we look at using test scores to analyse; (4) organisational structures and school choice; and (5) student bodies, under the umbrella of achieving equitable schooling.

5.1 Accountability mechanisms

Hamilton et al. (2002) define a test-based accountability system as "a set of policies and procedures that provide rewards or sanctions as a consequence of scores on large-scale achievement tests." The theory of change behind these

[8] We expand on the three areas focused on by the OECD in Woessman et al. (2007): accountability, autonomy, and choice.

systems is that public education can be improved simply by requiring all students to take standardised tests, as a kind of information intervention.[9] Schools are then punished or rewarded depending on their performance. This strategy is based on the typical principal-agent problem faced by the economist, where testing provides feedback to governments about schools' behaviour. In practicality, the role for tests using this definition is to ensure that students are achieving 'good' results (however that may be defined by the government). For Australia, this ties into the recent recommendation to "Enhance school and system internal self-review and external quality assurance processes for the purposes of monitoring and reviewing student learning gain" (Gonski et al., 2018).

A large body of literature argues that school accountability mechanisms, such as standardised testing, lead to improvements in outcomes (McDermott, 2011; Hamilton et al., 2002; Peterson and West, 2003). At the individual-level, it has been shown in the behavioural laboratory by Levitt et al. (2016) that when students are faced with both financial and non-financial incentives to perform better on tests, they do. While the literature agrees that increased accountability works in theory and on the individual-level, little has been done to evaluate the Australian system to see if it is fulfilling its potential - despite the fact that a by-product of NAPLAN testing is a vast source of data on educational outcomes.

The international literature, specifically in the United States where it has been a flagship policy, is quite detailed on testing in terms of an accountability mechanism and its effect on achievement. Hanushek and Raymond (2005) analyse the effect of the accountability mechanisms introduced in the 1990s on achieve-

[9] Information interventions need not only involve standardised testing, as will discussed later in this section. See Pandey et al. (2009) as an example of some other types of school-based information interventions aimed at improving educational outcomes in India. See Cornell-Farrow (2014) for more on accountability mechanisms in schooling and NAPLAN testing.

ment growth in various states as measured by the National Assessment of Educational Progress (NAEP), by looking at differences in achievement growth across states over time. As discussed previously, the United States enforced a policy whereby reform in schools was driven by high-stakes testing. Their analysis shows that these implementations had a clear positive impact on achievement over the period studied. However, it was also found that this reform had differential impacts on certain types of students. For example, over the period studied the black-white test score gap worsened.

A natural experiment in the United Kingdom, where league tables[10] were abolished in Wales but not England, tells a similar story of the heterogeneous impacts of accountability mechanisms on student achievement. Burgess et al. (2013) finds that the abolishment of league tables in Wales reduced school effectiveness when compared to England, and again that this policy affected schools heterogeneously. Most interestingly, removing the league tables had no impact on the performance of schools already performing in the top quartile. This may mean that league tables are effective at improving outcomes in average-to-lower performing schools.

This league table story is of particular interest to the Australian case, where the My School website behaves in a manner similar to league tables. The key difference is that, in Australia, schools can only be compared that are 'statistically similar' to each other, based on their ICSEA measures. As NAPLAN testing and the My School website were introduced in similar time periods, it is, however, difficult to undertake any analysis of the impacts this accountability mechanism has had on student and school performance. A lack of data from the previous periods render the possibility of a difference-in-difference style analysis

[10] A method whereby schools are publicly ranked on achievement.

impossible using the NAPLAN data. While testing existed in many states before the implementation of NAPLAN, there is little way of knowing if the scores achieved on these various state-wide tests are comparable to those in NAPLAN, complicating the researcher's ability to disentangle the effect that the publication of school-average NAPLAN scores online has had on overall achievement. It would also be difficult to know if any increases in student achievement are due to cheating. For example, Battistin et al. (2017) show clear incidence of cheating in Italian schools, thanks to a natural experiment where only some schools are randomly assigned a monitor. Once score manipulation on INVALSI testing is taken into account, regional rankings for academic performance are in fact reversed. This cheating may be due to both the real and reputational risks that are faced by schools that achieve low scores.

If we take cheating into account, this may tell a much different story as to why Hanushek and Raymond (2005) and Burgess et al. (2013) find that accountability mechanisms are affecting school performance differently. Perhaps, faced with the negative ramifications of low performance, schools are not actually improving their practices, instead partaking in cheating or manipulative behaviours. These need not be negative or unethical behaviours. For example, Reback (2008) has also shown that accountability systems may not just raise levels of student achievement, but change the *distribution* of student achievement. Looking at the United States where schools must meet a level of minimum competency, they show that this policy causes schools to reallocate their resources to focus on improving the performance of students who are on the pass/fail margin. Testing may therefore improve the scores of these children more, at the detriment of others. It is also difficult to know if these students are being coached to improve their short-run performance, or if they are actually reaching a higher level of educational performance in the long-run. It is also difficult to know if in response to tests, schools limit their curriculum to 'teach to the test', thus improving test scores but diminishing the quality of education students receive.[11]

Perhaps given the drawbacks associated with testing regimes, as discussed above, school systems around the globe are shifting to more nuanced and sophisticated accountability tools. This may include external school monitoring, such as the Office for Standards in Education (Ofsted) scheme in the United Kingdom where schools are regularly inspected, or self-evaluation tools that emphasise community engagement like the 'report card' systems in countries such as India and Brazil or the Ghanian process of School Performance Appraisal Meetings (Cornell-Farrow, 2014). It has been recommended that Australia move to an online version of NAPLAN,[12] followed by a proposed shift to an online formative assessment tool that is of greater benefit to schools, teachers, and thus students. This has particular ramifications for the role of NAPLAN data in policy-making and economic analysis. On one hand, it may cause the loss of a significant data source for governments to provide evidence-based policy solutions, but on the other it may improve the usefulness of the data; for example, the issues involved with the time lag of the data would be solved.[13] It could prove invaluable, providing that the system retains elements of collecting compulsory, standardised measures across cohorts. An example of this is the app Classroom Monitor, developed for monitoring classroom achievement in the UK, which

[11] Lazear (2006) provides a theory for how best to incentivise high-stakes testing regimes, with a comparison to deterrents for speeding.

[12] A process for which pilot studies have recently been undertaken with questions raised as to the comparability of data collected from written and online tests.

[13] Note that Levitt et al. (2016) also show that the power of incentives to motivate students to perform well on tests disappears if there is a delay between the test and receiving the reward.

also provides pre-set reporting tools for reporting to organisations such as Ofsted. A program such as this could allow schools to develop internal formative assessment tools within a framework that also allows for the collection of standardised information across schools for use by policy-makers. This would also increase student familiarity with using the tool, which could ultimately lead to better quality information.

In terms of the role of NAPLAN in economic decision-making, the move to an online-based formative tool could therefore in fact improve the kinds of analysis that can be done. For example, the tool would necessitate creating student accounts which would stay with the individual across their school life. The approach could thus increase the value of the NAPLAN data, instead of making it redundant as some posit.

5.2 Impacts of school funding

There are numerous arguments for the public funding of schools - whether this be a positive externality argument, an ethical argument for equality among citizens, or another. In Australia, Parish (1963) argued that “it is the proper function of government to encourage investment in education,” as well as that government funding in the education sector should extend to all types of schools, including the private system. Given that governments are providing billions of dollars to schools each year, it is natural that an analysis of the efficiency of these funds be undertaken. This plays into the theory of accountability mechanisms detailed above, where schools must justify that they are spending money effectively by providing a good quality education to their students.

Levin (1974) and Hanushek (1979) discuss the educational production function, which considers schools to behave like a type of firm producing the output ‘education’. This output is generally measured as school-average scores in standardised tests. Output is determined by two inputs - school resources and average student background. This theoretical framework has informed econometric specifications for analysing school funding in the literature.

Internationally, the literature has been divided as to how resources impact student test score achievement. In the United States, Card and Payne (2002) studied micro-samples of SAT scores (a standardised test for college admission) using a series of models to look at how school finance reform and redistributions of school spending affect the distribution of test scores. It was found that test score gaps between disadvantaged communities can be narrowed by equalising spending across richer and poorer districts. In the state of Maine, Deller and Rudnicki (1993) constructed a set of nominally efficient schools, or schools that are able to maximise student achievement given their particular characteristics. This study identified inefficiencies in production in these schools, but found it difficult to pinpoint a particular policy pattern behind their result. Graddy and Stevens (2005) have also found different results for the efficiency of school resources across various school types. By studying panel data from the Independent Schools Information Service in the United Kingdom, it was found that decreased student-teacher ratios (*i.e.*, hiring more teachers) are related to higher examination results. This result is interesting given that this has not been found to be the case in any state schools in the United Kingdom.

What has been highlighted is that the results of analysis of school funding or resources are highly dependent on context, and often lead to conflicting conclusions. Eric A. Hanushek, Paul and Jean Hanna Senior Fellow at Stanford’s Hoover Institution, has written prolifically on the issues involved with the estimation of educational production functions. Hanushek (1979) details the lack of conceptual clarity and analytical problems involved with the estimation of educational production functions. The fact that this literature crosses many disciplinary boundaries; for example, that of the education literature, the public policy literature, the social sciences literature, and the economics literature, may also be a reason behind the conflicting results of analyses thus far. Hanushek discusses that the original use of input-output analysis in the Coleman Report[14] morphed into the economist's educational production function - which does not come without ramifications regarding the interpretation of results. The major difference between an input-output analysis and the production function is that the latter absorbs the concept of *maximisation*, that is, that a school is performing efficiently in turning resources into educational output. Hanushek argues that no one would expect a firm to change its behaviour given the estimation of a theoretical production function, so why should these be taken so seriously in the case of schooling? A number of conceptual limitations to the estimation of these functions are outlined:

- Standard production functions have some kind of homogeneous output produced at various numbers of unit. What is a unit of schooling?

- In fact, most schools produce *multiple* outputs. Standardised test scores represent just one measure for these possible outputs.

- There are relatively fixed levels of labour and capital in a school, *i.e.*, one teacher with a number of students in their class that may vary slightly. This lack of variation tends to mean that it explains very little in terms of the variation in outcomes.

[14] A study undertaken in the United States of America on the equality of educational opportunity, mandated by the Civil Rights Act of 1964.

- Choices in inputs tend to be guided by data availability rather than any conceptual notions. Dewey et al. (2000) further discuss how the inclusion of parental income in these functions can confound demand and production, leading to misspecification.

- It is nigh impossible to decipher if schools are behaving efficiently in production.

Other issues include deciding on a functional form, the level of aggregation (Hanushek et al., 1996), accounting for selection effects into certain schools, multicollinearity, addressing assumptions such as that ability remains constant over time (Ding and Lehrer, 2014), including the cumulative nature of schooling (Todd and Wolpin, 2003), not to mention choosing statistical methods correctly. Hanushek (2003) bemoans the lack of understanding of incentives at play in input-based analyses of schooling, as well as exploring the variations in teacher quality not affected by resources that affect performance.[15]

While it is exceedingly difficult to estimate the true impact of resources in schools, this should not detract from attempts to quantify them as much as possible. In Australia, the literature has been particularly focused on this issue due to its political pre-eminence in educational policy discourse. A concerted effort to understand how funding is working in Australian schools is key to the cogency of this debate. Research thus far has highlighted the inefficiency of the funding of schools in Australia. Watson and Ryan (2010) argue that as private schools typically use extra government funding to hire more teachers rather than to decrease their school fees charged to parents, noted as early as

[15]See Hanushek and Rivkin (2010); Hanushek (2011); Chetty, Friedman and Rockoff (2014*a*,*b*) for further analysis of the economic analysis of the impacts of teachers. This literature is outside the scope of possibilities for the NAPLAN data at this point in time.

Williams (1985), current funding schemes undermine the public system and fail to improve access to the private system for those from lower socio-economic backgrounds. In Victoria, Dancer and Blackburn (2017) find only 37.5% of school studied are efficient. Similarly, Blackburn et al. (2014) find schools in New South Wales to be moderately inefficient. Although it is again highlighted that context matters, as those with the most favourable socio-economic environments are most efficient. They also note the efficiency gains from increased enrolments at a school.[16] In Victoria, Cobb-Clark and Jha (2013) estimate the relationship of per-pupil expenditure on achievement in test scores using a value-added model accounting for lagged achievement, and find that these two variables are only moderately related. Finally, an input-output analysis undertaken by Nghiem et al. (2016) find that NAPLAN test score growth could be improved by 64%, on average, by learning from best practice.

The My School website phenomenon has also meant the collection of financial data alongside test score data, allowing a broader view on the efficiency of government spending in schools across all states and territories for the first time. Previous attempts at analysis have focused on much smaller sample sizes, in only particular school systems or states. As outlined above, most analyses of school funding cannot be easily generalised outside of their particular context. The NAPLAN data therefore provide an exciting opportunity to shed light on funding across Australia as a whole, which is of particular use to the federal government as they are increasingly spending more in schools. This is also useful for state governments to make better-informed policy comparisons between

[16]Note that Chakraborty and Harper (2017) contradict this result, finding that schools in New South Wales are quite efficient using a technical inefficiency effects model. They find that primary schools are 88.6% efficient and secondary schools 96.4% efficient. However, this paper is more focused on the impacts of socio-economic factors on school efficiency, than overall resource efficiency.

each other, to encourage best practice within the federal system. While the financial data is limited at this point in time, for example; it does not provide detailed breakdowns on how funding is spent within a school, it provides a useful starting point from which to reposition the funding debate towards evidence.

5.3 School management

In their analysis of school efficiency, Cobb-Clark and Jha (2013) also find that budget allocation matters when looking at how schools produce outcomes. Santin and Sicilia (2015) similarly conclude that if the responsibility for school budget distribution is given to principals themselves, there are positive effects for the efficiency of public schools in Uruguay. School leadership and management clearly play an important role in the ability of a school to produce higher outcomes. A further recommendation of the second Gonski review is to empower school leaders, as well as to support a profession of expert educators. The following section will therefore explore how the NAPLAN data can be used to better understand best practice for school management.

In Brazil, Tavares (2015) finds that management practices such as performance monitoring and target-setting led to improvements in math scores, especially in disadvantaged schools. These improvements are most likely due to changes in management and teaching practice. In an analysis of 1800 high schools in 8 countries, Bloom et al. (2015) also find that schools with higher quality management achieve better educational outcomes. In a slightly tangential study, focused on the private schooling sector, Green et al. (2011) conclude that private schools in the United Kingdom have been most successful at translating their curriculum into producing the outputs demanded by the economy. The management direction of a school clearly plays a strong role in determining how successful that school will be.

The question must be therefore be asked as to if and how NAPLAN can contribute to the understanding of school management, and if this data can play a role in constructing, for example, a management index to isolate best practice. Ultimately, the managerial quality of a school is what one is trying to determine through the implementation of the various accountability mechanisms discussed in the previous sections. One goal of NAPLAN testing, and the My School website, would be to incentivise school management to change their practices and innovate in such a way that will improve their educational outcomes.

Coelli et al. (2018) undertook a survey of school principals in Australia, to understand if school managers responded to the increased public scrutiny of NAPLAN test score reporting on the My School website. This study attempts to look inside the 'black box' of the educational production function, to uncover what policies and practices schools have actually implemented over the period of study. Principals from all school sectors were surveyed, both prior to and then three years following the publication of the first round of data online. The first survey uncovered differences in policy and practice between low-performing schools and high-performing schools, as well as between school sectors. For example, low-performing schools are most likely to have shorter school days, and less involved parents; however, are more likely to use interventions such as tutoring of struggling students. Private schools were most likely to set more homework, as well as to incentivise teachers to produce high outcomes, such as with the threat of dismissal. Their results show that the My School website did not cause already low-performing schools to change their practices. There was, however, evidence of narrowing the curriculum and directing resources to the subject areas covered in NAPLAN, to the detriment of other areas. One potential reason for this could be that education policy is so rigid that these principals do not have any real managerial power over how their schools are run. It may also be the case that they do not prioritise My School publications, or that there has not been enough time for the policy to make a real long-term impact.

What this study shows is that while the NAPLAN data is important to uncover which schools are performing poorly, it does not say enough about what policies and practices are occurring in schools to truly diagnose these problems. The data alone is not enough to determine which schools are managed best, or which schools are behaving most efficiently. What the data could do is target particular schools (using proper random sampling techniques, and stratifying by levels of achievement) on which further data could then be collected to uncover key differences between low-performing and high-performing schools of similar backgrounds. This, however, would require significant commitment from government and the education sector at large to allow this to occur.

5.4 School choice

School choice, particularly between the public and private schooling sectors, has been an especially popular topic in the economic literature on schooling. This is likely due to the role it plays in government decision-making about which sectors to allocate public funding to, and at what levels. This is also an important topic for parents when making decisions about where to send their children to school. Standardised test scores are the measure that is most often used to compare performance between schools, and school sectors.[17]

[17] Note that there is also a literature on how school sector impacts later life outcomes, for example; Polidano and Jha (2015) find that wages for students from Catholic schools grow at a faster rate with labour market experience than those of a public school graduate. They also find a similar result for those from independent private schools, arguing that these school sectors may be better at preparing students for work.

The literature is contentious as to whether private schools cause students to perform better. At first glance, it would appear that private schools score higher on tests. However, it is difficult to disentangle this effect from a self-selection effect. Hsieh and Urquiola (2006) find that a voucher program in Chile, which enabled any student to attend private school if they so wished, increased sorting as the 'best' students left for the private sector; however, there was no effect on test scores. It is arguable that both parents with academically gifted students as well as those who most value education will be willing to pay to send their child to a private school. This self-selection effect is what causes the improved test scores in private schools, not the education provided. Indeed, there is a large part of the literature focused on how best to combat the problems involved with estimating the effect of private schooling, such as Altonji et al. (2000). This study proposes a method by which to use the selection on observables to guide the amount of selection on unobservables involved in school choice.[18] This literature has wider ramifications for approaches to estimation with selection on unobservables in applied microeconomics at large.

In terms of the findings of the literature, the case has been argued for both sides. In England, Gibbons and Silva (2011) find that students in religious primary schools progress faster than those in the public system, but that this is explained away by self-selection on pre-existing characteristics. A similar result is found by Elder and Jepsen (2014) in the United States and Chudgar and Quin (2012) in India. In Australia, Ryan (2013) find that declines in PISA scores are more evident in private schools than in public schools. In contrast, Lefebvre et al. (2011) find that the percentile rank of a student in math will increase by between 4 and 10 points for those switching from a public to a private

[18] See Cardak and Vecci (2013) for use of a similar methodology in Australia.

school (the treatment effect on the treated). In their analysis of charter schools, Zimmer et al. (2012) highlight that the assumptions of the modelling approach play a very real role when estimating the impacts of schools, and that this may be why the literature has shown mixed results thus far.

Another explanation for why students may perform better in private schools is that of a 'peer effect'. That is, students will perform better when they are in a better environment with high-performing fellow students. Using a subset of the NAPLAN data from the Victorian Government, McVicar et al. (2018) find that there are positive effects from having higher-achieving peers. However, in their analysis of exam schools in Boston and New York, Abdulkadiroglu et al. (2014) find that peer characteristics do not have a clear causal effect on test scores. By exploiting the cutoff scores for attendance at these schools, they are able to implement a fuzzy regression discontinuity design to compare students above and below the threshold who attended schools with different peers. Ellison and Swanson (2016) also find a lack of evidence for the direct impact of peer effects, through a non-parametric analysis of scores achieved in the American Mathematics Competitions. Ryan (2017*a*) explains the difficulties of estimating peer effects, and argues that the methods used in the literature are often contentious if they are not based on some source of external manipulation of peer groups. This may explain the differing results in this body of literature.

Nevertheless, it remains clear that schools produce high-achieving students at different rates. Ellison and Swanson (2016) illustrate that there are differences in the frequency between seemingly-similar schools, yet struggle to provide an explanation for this. Perhaps the public/private debate is too broad, but school choice at the micro-level does matter. In a survey of the literature, Bast and Walberg (2004) find that parents are just as good at rating the performance of schools as experts, and that students who attend their 'school of choice' will perform better than those in their assigned public schools. This supports the finding in Chicago of Cullen et al. (2005) that students who are assigned a school and then opt out to attend a different school do better than those who stay in that to which they were assigned.[19] The ability of parents to best match their child to a school may explain the success of the charter school movement in the United States.[20] For example, Hanushek et al. (2007) find that the parental decision to leave a charter school is significantly related to quality. While charter schools, on average, do not perform better than public schools, there is considerable heterogeneity in the sector. By reducing the transaction costs for students to switch schools, it may incentivise schools to improve their quality to attract students.[21]

The effect of private schooling is also contested in the Australian literature. One must also note that the findings in other countries cannot be directly related to the Australian context, as we have a particularly large and heterogeneous private sector, where not all schools are necessarily expensive and 'elite'. Furthermore, parents choose private schooling for reasons that may differ to those by parents in other countries.[22]

In their analysis of the early NAPLAN data, Miller and Voon (2012) find that private schools achieve higher test scores on average than public schools,

[19]This feeds into the literature on school vouchers. See Levin (1998) for a discussion of the effectiveness and costs of voucher programs.

[20]Note that school choice policy mechanisms could include affirmative action policies, see Doan (2016).

[21]Note that this must also be understood in terms of a capacity approach, where schools can only improve given the resources to do so, not just through incentives. This forms part of the analysis of school funding in previous sections, as well as an understanding of a school's student body, which will be explored in the following section.

[22]See Dearden et al. (2011) for a comparison of the determinants of school choice between the United Kingdom and Australia.

but that this effect may be due to selection processes. Nghiem et al. (2015) use a subset of the NAPLAN data, linked to data from the Longitudinal Study of Australian Children (LSAC), and find that school type does not affect either cognitive or non-cognitive development in children.

The expanded individual-level NAPLAN data set provides an opportunity to analyse the effect of school choice on outcomes in more detail. With background information on observable characteristics, and information about the background and test scores of a student's peers, we may be able to better isolate the effect of switching school sectors, or even the effect of particular individual schools. This could be a powerful tool not only for governments, but also for schools who want to understand how best to serve their students. This, of course, would come along with a number of ethical issues and would need to be conducted sensitively. At the level of public versus private schooling, we can begin to get an idea of the effect of school sector by using observables such as the work and education status of the students' parents. However, this area of analysis would benefit from a deeper study using Census data about what determines a parent's decision to send their child to a private school.

This area of study is also currently restricted by the inability to follow a student between schools through time. If we could follow students throughout time, and observe them switching between systems, we would be able to perform analysis on the switchers such as that undertaken by Lefebvre et al. (2011). We would argue that this would be the best methodology for clearly identifying the effect of private schooling in Australia. A second approach would be to implement a matching approach; however, as the background characteristics contained in the data are limited at present, it would be difficult to ensure that we are truly matching similar students across sectors. It would therefore require further collection of more detailed background characteristic data to best implement a matching approach.

5.5 The provision of equitable schooling

The final question that must be asked, and indeed has been as early as Summers and Wolfe (1977), is if schools really can make a difference at all. A primary goal of the Australian federal government, and indeed perhaps the driving force of both Gonski reviews themselves, is that of “Equipping **every** student to grow and succeed in a changing world.”[23] It is, however, disputed if schools are able to change or mould outcomes for students given their pre-existing characteristics and home life. To take the education production function approach, are the only inputs that truly matter those of the student themselves? For example, Polidano et al. (2013) find that the lower educational aspirations of low SES students and their parents is one of the most important contributing factors of lower achievement in Australia, not differences in school characteristics. This is a particularly popular area for economic analysis, as will be detailed presently.

Cullen et al. (2013) have argued that high schools typically have student bodies that have outside options to schooling, as well as exhibiting great heterogeneity, making it difficult for schools to prepare their students for the world. They further note that this is particularly the case for high schools with disadvantaged students or in disadvantaged neighbourhoods. While their conclusion is that these schools should realign their goals to focus on practical life and labour market skills, instead of the traditional focus on testing, their study draws particular attention to the question of how schools can cater to students with such vastly different backgrounds and needs, in a world focused on providing

[23]Emphasis my own.

equitable education for all. Our task in this section is to analyse the literature on disadvantage so far, and determine how and if test scores really can play a role in understanding these issues.

A popular area for analysis, particularly in the United States, has been to analyse the impacts of racial background on achievement. Roland G. Fryer, Jr. and Steven D. Levitt have studied the question of the black-white test score gap in depth, both in the first two years of school (Fryer and Levitt, 2004) and through the third grade (Fryer and Levitt, 2005). Using data from the Early Childhood Longitudinal Survey, they show that, once controlling for a small number of background covariates, there is no gap in test score achievement between students from white and black backgrounds. However, they find that over the first two years of school, black students typically fall behind faster than their peers from other racial backgrounds (Fryer and Levitt, 2004). By the third grade, the test score gap can no longer be explained by observable background characteristics (Fryer and Levitt, 2005). The only explanation that can be suggested with some level of empirical backing is that it is something to do with school quality that is causing these differences (Fryer and Levitt, 2004), but this is disputed in Fryer and Levitt (2005).

Largely in Europe, the discussion of race has expanded to analyse how changing concentrations of immigrant students in the classroom may affect outcomes for both native-born and immigrant students. In Denmark, using PISA data, Jensen and Rasmussen (2011) find that by increasing the concentration of immigrant students in a classroom, students achieve lower scores in both reading and math, with a stronger effect for native students. This has also been studied in Italy, with Ballatore et al. (2018) identifying a pure ethnic composition effect by exploiting the rules of class formation governed by Maimonides' rule.[24] These kinds of rules for governing class size specify a particular number of students at which the class will be split in half. In Italy it is difficult to predict if classes will reach the threshold at which classes are split, creating an element of randomisation in class sizes. It also creates randomisation as to whether new immigrants are 'added' to a class (the class is not split), or whether they 'replace' native students in a class (the class is split). By analysing INVALSI test data, they find that there is a negative effect on native test scores, especially in the case of first-generation immigrants. This may be due to disruptions in class due to language barriers.

The literature shows that it may not only be the composition in the classroom that matters for test score outcomes, but even the composition of your neighbourhood. While Gibbons et al. (2013) find that changes over time in neighbourhood composition do not affect test scores in England, Nicoletti and Rabe (2013) find that neighbourhood may explain up to 10-15% of the variance in test score performance. Nevertheless, family plays a much larger role in explaining the gap, with family background explaining 44-55% of the variance in test scores. The reasons that family background may influence your school performance are varied. Behrman and Knowles (1999) illustrate the considerable relationship between household income and a child's educational success in Vietnam. Stevens and Schaller (2011) show that, in the short-run, a child will suffer from educational difficulties when faced with a parent losing their job. Using an experimental approach in schools in Chicago Heights, Fryer et al. (2015) show that a student's test scores can be improved by incentivising par-

[24]Maimonides was a Jewish scholar in the Middle Ages, who first noticed a correlation between class size and student achievement.

ents to participate in their child's education. While the literature makes it clear that family characteristics have a significant impact on educational outcomes, this remains a significant limitation of the NAPLAN data at present. While both the educational literature (Ford, 2013) and the economic literature (Cobb-Clark and Nguyen, 2012) have highlighted that a detailed understanding of racial, economic, language, educational, and social backgrounds of a child's family will determine their educational outcomes, these variables are only included at broad levels of granularity. For example, there are only four categories to cover all possible jobs that a parent may have; while we know that a student comes from a language background other than English, we don't know what that language is; we have no understanding of a child's living situation; and we know little about their parent(s)' income. While we can understand these questions in broad strokes with the current data set, these are important contextual questions that must be clarified in greater detail before we can truly begin to understand how policy changes may differentially affect individual students.

There are also individual-level characteristics that may affect a student's performance, specifically their gender and their age. It is generally conventional wisdom, and supported by much of the literature, that girls have an advantage in reading, and boys have an advantage in math. Bedard and Cho (2010) analyse the gender test-score gap across a number of OECD countries using the TIMSS data. However, countries with pro-female sorting or policies in place have smaller observed gender gaps in math than those that did not. Moreover, the gender gap can be influenced by regional characteristics. For example, using micro-data from SACMEQ in 19 African countries, Dickerson et al. (2015) find that fertility levels in a region are a determinant of the size of the gender gap in mathematics in Africa. In countries such as Pakistan, girls may be disadvantaged in all subject areas. Using surveys of rural households, Alderman et al. (1996) find that girls are at a significant educational disadvantage in developing countries, and that both demand and supply side factors may help to improve this gap. The gender gap may also change over time, as with the black-white test score gap. Fryer and Levitt (2010) indicate that in the United States there is no difference in math performance between boys and girls when they start school, but that over time girls begin to fall behind. Questions related to gender can be studied well with the existing NAPLAN data. Gender is included as a covariate for each individual student, and we can therefore also deduce from the genders of their classmates if they are attending a same-sex or co-educational school. Due to the large sample size, the Australian data may have significant promise for analysing the impacts of same-sex schooling on outcomes, for both genders.

Finally, a large part of the literature focuses on the age of starting school, as this is a policy that can be directly controlled by governments. Using two birth cohorts in the United Kingdom, Crawford et al. (2014) find that children born earlier in the year do better in tests, and that age adjustments to scores may be necessary. Black et al. (2011) also find that older children in Norway perform better on tests, and that IQ scores are slightly negatively affected by starting school later.[25] However, while older children do better this does not mean that it is necessarily of benefit to hold back your child from starting school. Fletcher and Kim (2016) find that starting kindergarten earlier increases average test scores in the United States. It has also been found in Australia, using the LSAC subset of data, that starting school earlier improved cognitive skills, particularly

[25] In Norway, there is a discontinuity around 1 January as to whether students start school in that calendar year or not.

for socio-economically disadvantaged students (Suziedelyte and Zhu, 2015). In analysing the life-cycle effects of school starting age in Sweden, Fredriksson and Ockert (2014) argue that the literature finding that older students perform better is confounded, and that there may be differing effects over the life-cycle but no absolute difference between older and younger students. While student ages are built into the individual-level NAPLAN data set, it remains more difficult to analyse these kinds of questions in the Australian context.[26] This is due to differences in policies about the age of starting school across states. For example, Queensland only introduced a compulsory full year of prep (for students aged 5, also known as reception) in 2017, bringing them into line with the thirteen years of schooling usual in the other states.[27] There are also differences in rules as to when students start their first year of school, with some doing three, four, or five terms of reception. We would require data on at what age a student started school to be able to analyse these kinds of questions. At this point in time, however, it could be possible to analyse differences between how well older and younger students within the same grade level perform on NAPLAN testing.

6 Concluding remarks

This paper has explored the various areas in which NAPLAN data may be useful for economic analysis and for policy-makers. Particularly, we believe that there is much to learn about the following topics, given the NAPLAN data in its current form at the time of writing. Firstly,

[26] Note that these questions may be better suited to the LSAC subset of the NAPLAN data, such as in Suziedelyte and Zhu (2015).

[27] Miller and Voon (2014) illustrate the lower outcomes achieved by schools in Queensland using a regression discontinuity approach comparing this state to New South Wales.

(1) Use the financial information contained in the My School set of data, that is, the school-average level data, to develop a greater understanding of how funding is really impacting Australian schools (both public and private);

(2) Use the student-level data to examine what kinds of students are at risk of performing poorly on tests such as NAPLAN, in order to understand where policy can be best targeted.

Item (1) is of importance following from our discussions of accountability mechanisms and the efficiency of school funding. The concept of school choice is also closely linked to an understanding of how funding is best utilised, as governments decide where to allocate their funding. Item (2) is also a pivotal piece of understanding to create accountability mechanisms that are fair, as this will only be made possible through a thorough understanding of how the goal of equitable schooling plays into these policy decisions.

It is worth noting that more can be done to improve the econometric uses of the data, particularly the implementation of a unique student identifier as detailed in the Gonski report. The collection of more detailed background statistics, and for these to be made available to the researcher given ethics clearance, would also vastly improve the quality of analysis and thus the evidence base from which policy can be determined. Despite these drawbacks, which do significantly affect the econometrician's ability to achieve identification and detect causality in the field of education, the results of analysis of this kind using the NAPLAN data must not be underestimated. This data still has much to tell about the nature of schooling in Australia.

THE UNIVERSITY of ADELAIDE

Statement of Authorship

School funding and student achievement on standardised tests: Empirical evidence from Australia
Unpublished and unsubmitted work written in manuscript style

Principal Author

Name: Sarah Cornell-Farrow
Contribution: Acquisition and cleaning of data. Performed estimations and analysis, interpreted results, and wrote manuscript.
Overall percentage (%): 70%
Certification: This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

25 January 2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate to include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name: Firmin Doko Tchatoka
Contribution: Parts of methodology and manuscript editing.

25 January 2019

Name: Virginie Masson
Contribution: Parts of introduction, results interpretation, and manuscript editing.

25 January 2019

School funding and student achievement on standardised tests: Empirical evidence from Australia*

Sarah Cornell-Farrow, Firmin Doko Tchatoka[†] and Virginie Masson

The University of Adelaide, School of Economics, Adelaide, SA, Australia, 5005.

Abstract

This study sheds new light on the relationship between school funding and student achievement on standardised tests in Australian primary schools. We use a school-average test score data set from the National Assessment Program - Literacy and Numeracy (NAPLAN) paired with funding data for each school to determine how the three different types of funding received by Australian schools impact test scores differently depending on sector and state. We find, in general, that funding from the federal government has the least beneficial impact, with state government funding and parent fees more likely to provide the greatest benefit to schools. The policy implication of these results is that federalism has a role to play in delivering successful funding policy to Australian schools, and that funding is most beneficial when provided at as local a level as possible.

Keywords: National testing, school funding, quantile regression, monte carlo.
JEL Classification: C13, C21, C23, H52, I21, I29.

1 Introduction

Policy-makers and school officials often argue for increased funding to poorly performing schools in order to improve their learning outcomes. In turn, test scores provide regular and consistent measurement of these outcomes, and thus a systematic way to allocate funds on the basis of performance. As a result, school

*This research is supported by an Australian Government Research Training Program (RTP) Scholarship. Thanks must also be given to the Australian Curriculum, Assessment and Reporting Authority for the provision of the data utilised by this study. Thank you to participants at the 2016 Australasian Meeting of the Econometric Society, and a number of anonymous referees for useful comments that have improved this paper.

[†]Corresponding author: firmin.dokotchatoka@adelaide.edu.au

accountability mechanisms are increasingly linking monetary incentives to performance on standardised tests. Under the United States' *No Child Left Behind* policy, for example, schools not achieving certain pass rates can be faced with monetary sanctions. In a contrasting strategy, the Australian *Review of Funding for Schooling* advocated the targeting of *increased* funding to schools achieving low scores on standardised tests. There is, however, little empirical evidence that higher levels of funding will lead to higher levels of achievement.[1] In fact, while funding in Australian schools has increased each year, performance in international test programs such as PISA has declined. Funding increases nevertheless continue to be the major focus of Australian education policy. We therefore aim to understand if there is a causal relationship between funding and test scores in Australian primary schools, using data from the National Assessment Program - Literacy and Numeracy (NAPLAN).

The impacts of school resources on educational outcomes are notoriously difficult to estimate. We argue that this difficulty is the result of heterogeneity in impacts across the test score distribution.[2] Schools that are already performing well will be little impacted by increased levels of funding, and schools performing poorly may well be restricted by their capacity, or some kind of unknown school fixed effect. To compound the problem, schools achieving high scores are often high fee private schools. While some studies may argue that high levels of resources in these schools lead to improved educational outcomes, the students in these schools are in fact a selected population with above-average abilities or background characteristics.[3] The higher fees in these schools are likely an

[1] See Hanushek (2003) and Cullen et al. (2013) for discussions on the failure of resource increases to improve educational outcomes.

[2] See Bitler et al. (2006) for similar arguments regarding the effects of welfare reform and Powell (2014) for the effects of economic stimulus payments across the wealth distribution.

[3] See Levin (1998).

indicator of their desirability for high-performing students, rather than driving the improved results themselves. We therefore propose a method to estimate the effect of funding in Australian primary schools on reading and numeracy test scores, taking into account heterogeneity in effects over the score distribution. Additionally, we account for potential reverse causality between funding and scores to allow for the possibility that scores obtained in preceding years are determining the level of fees charged by a school.

It is generally well accepted that re-allocating resources within a school or school district can improve performance,[4] but the actual marginal effect of a dollar has not been clearly identified. In the United Kingdom, Green et al. (2011) have concluded that higher levels of resources in private schools contribute to improved educational attainment. Furthermore, Graddy and Stevens (2005) show that the lower teacher-student ratios in private schools lead to better outcomes for their students. In Australia, Miller and Voon (2012) found that average NAPLAN scores in private schools are consistently higher than those in public schools; however, hypothesise that these differences are explained by self-selection processes, rather than the increased levels of resources in private schools. To further this argument, Nghiem et al. (2015) found that, once controlling for sample selection, there is little difference in the outcomes produced by public and private schools. These findings in Australia are also supported by Elder and Jepsen (2014) in the United States and Gibbons and Silva (2011) in the United Kingdom. In contrast, Lefebvre et al. (2011) found that switching to

[4]In the United States, Card and Payne (2002) found that changes in the allocation of funding within school districts led to equalisation in the test score gap between disadvantaged and advantaged students. At the school-level, Cobb-Clark and Jha (2013) found that the strategic budget allocation of principals in Victorian public schools (who operate with a significant level of autonomy in financial decision-making) matters for improving student achievement. Furthermore, Reback (2008) has shown the strategic shifting of resources in Texan schools to help particular students improve their scores.

a private school in Canada improves a student's mathematics score by between 4 and 10 points. Card and Krueger (1994) also find that increased school resources are associated with higher educational attainment, although the range of their coefficients is wide. The amalgamation of test score data and funding information on the Australian *My School* website allows a unique opportunity to identify the effect of funding itself, in both public and private schools.

Specifically, using this data on test scores and school funding,[5] we investigate the effect of three major sources of funding - the federal government, the state government and parent fees - on school-average test scores in reading and numeracy. To our knowledge, this is the first paper to undertake a systematic analysis of the impacts of funding in Australian primary schools. Using an unconditional panel quantile regression approach that accounts for both unobserved school-fixed effects and endogeneity, we show that the relationship between funding and NAPLAN test scores differs between public and private schools, as well as varying along the distribution of test scores and across states. These are results that may be overlooked if implementing traditional mean-focused approaches such as fixed effects models. In particular, we show that federal funding generally has a negative impact on average test scores in both public and private schools. Meanwhile, state funding is in many cases positively correlated with average scores, especially in public schools in Victoria and private schools in New South Wales and South Australia. Our results also indicate that parent fees impact positively on average NAPLAN test scores for many schools.

The remainder of this paper is organised as follows. Section 2 describes the institutional context followed by a discussion of the data in Section 3. The

[5]Data were obtained from the Australian Curriculum, Assessment and Reporting Authority (ACARA).

methodology is discussed in Section 4 and results are presented in Section 5. Section 6 provides some discussion and concluding remarks.

2 Institutional context

2.1 The Australian school system

There are three sectors of schools in Australia, with the largest being the public system, which is administered and funded by the government. These schools provide a secular education to students living in the local area at a minimal cost, although a minority of these schools are selective on other characteristics such as a student's ability. The second largest schooling sector is the Catholic sector. These schools provide a Catholic education to those of the Catholic faith, and are also mainly funded by the government. They, however, differ from public schools in that they provide an education based on a Christian ethos.[6] Lastly, the 'private' sector is made up of non-government or independent institutions that have their own governing board responsible for the school's operation. These schools often provide religious education, *e.g.*, Anglican, Methodist, and are more expensive to attend and traditional in style. However, some independent schools simply provide different curriculums such as the Steiner system of independent schools, or are geared towards special interests such as sport or music programs. While these schools are governed independently, they still receive government funding. In addition, these schools charge parent fees that vary significantly from levels similar to those charged by public schools to thousands of dollars a year, and many private schools provide scholarships to attract talented

[6]We do not consider Catholic schools as they are mostly funded by the government yet differ in the type of education they provide.

students without the means to attend.

2.2 Policy context

Australia has a federal style of government, combining a national government (often referred to as the 'Commonwealth' or the 'federal' government) with state and territory governments, and local governments. Constitutional responsibility for the delivery of school education rests with the state and territory governments; however, these governments co-operate with the Australian government via the Ministerial Council for Education, Early Childhood Development, and Youth Affairs (MCEEDYA)[7] to develop and co-ordinate goals for improving educational outcomes and strategies for achieving them nationally.

Federalism allows for variance in government policy throughout the country, in order to better accommodate for a diverse citizenry, as well as the customisation of policy to best meet local community needs. Theoretically, this variance causes competition between states and territories, which becomes an incentive to provide the best services possible. Nevertheless, this system comes with many potential disadvantages, notably in the realm of education policy delivery. While the state and territory governments hold residual Constitutional power for the provision of education under section 197, the Commonwealth has increasingly encroached on this policy area through the use of section 96, which allows the federal government to provide grants on such terms as the parliament sees fit. Furthermore, state and territory governments are restrained financially due to vertical fiscal imbalance, where the Commonwealth is responsible for the majority of tax collection yet the states and territories are responsible for the

[7]This body has been replaced by the Council of Australian Government's Standing Council on School Education and Early Childhood since the period studied in this paper.

majority of public service provision. Therefore, in practice, responsibility for education funding is shared due to the federal government's power to make tied grants to the states. This can lead to inefficiency, duplication of funding and services, lack of transparency, and blame-shifting.

Different federal governments have used this responsibility in various ways, due to their differing ideologies in regards to the role of federalism in the political process. There was little federal involvement in school funding until the Karmel Report on schooling in 1973 established Commonwealth intervention in this policy area. At this time, the Whitlam Labor government was extensively increasing the number of grants provided to the states on the basis of creating the welfare state. The moves of the Whitlam government highlight that, in a federal system, the Commonwealth has a role to play in ensuring national standards are met across all states and territories. This is reflected in the longstanding commitment of all federal governments since the 1970s to provide a minimum grant to all schools, including non-government schools, to make sure that all students receive some form of government support. It was also found, and widely accepted at the time, that the funding of non-government schools has the extra benefit of taking pressure off the public sector to assume the costs of educating these students.

Despite the apparent consensus on the funding of schools, there remains contention in the amounts of funding and the levels of government from which this funding should come. While Whitlam saw a role for the Commonwealth to provide national leadership on welfare within the constraints of the federal system, the Hawke and Keating governments pragmatically saw federal reform as an opportunity to more efficiently use government resources. However, in an attempt to make the political system more efficient, this reform may have in fact increased inefficiency due to duplication.

Throughout the period of our study, the belief that the Commonwealth should provide national leadership was mirrored by the rhetoric of the 'Education Revolution' of the Rudd and Gillard Labor governments. This period saw the creation of the Australian Curriculum, Assessment and Reporting Authority (ACARA), the National Assessment Program - Literacy and Numeracy (NAPLAN), the *My School* website, and a National Curriculum.

In 2011, the government commissioned the *Review of Funding for Schooling*, popularly known as the Gonski report, the most comprehensive report of school funding since the 1973 Karmel report. The report explains that Australia needs effective funding arrangements for funding schools across all levels of government - arrangements that ensure resources are being provided where they are needed. It is argued that the existing arrangements are unnecessarily complex, lack coherence and transparency, and involve a duplication of funding effort in some areas. It was found that, due to current policy processes with their roots in the reforms of the 1970s, government funding of schools is marked by a distinct lack of co-ordination. Moreover, it is difficult to hold the government accountable for schooling, as it is unclear what level of government money is coming from, and what role each government should be playing. The Gonski review, however, did not establish a causal link between levels of funding and educational outcomes. Thus, this study has an important role to play in the context of this report, and in future reform of education funding policy.

2.3 School funding 2009-2012

In the time period studied by this paper, the roles and responsibilities of the federal government, and the state and territory governments, regarding the provision of school funding were described by the National Education Agreement (NEA). Under this agreement, the federal government assumed responsibility for providing funding to the state and territory governments, as well as non-government systems and schools. This was aimed at enabling schools to deliver better services to reach certain nationally-agreed outcomes and objectives. This agreement included \$42.4 billion[8] of funding to schools. Alongside this funding, the federal government signed national partnership agreements with state and territory governments to provide extra funding for certain areas. National partnership agreements in this time period included: Improving Teacher Quality (\$550 million), Education in Low Socio-economic Status School Communities (over \$1 billion), and Literacy and Numeracy (\$500 million).

A significant driver of this extra federal funding provided to schools just before our period of study was the Global Financial Crisis, with the funding forming part of the fiscal stimulus programme of the Australian government. Particularly controversial was the provision of one-off extra capital funding to schools for the building of general purpose and library buildings. It must be noted that, for this reason, we do not consider capital expenditure in this paper. Instead, we focus on recurrent funding that is provided yearly to schools.

2.3.1 Public schools

Public schools are funded by their relevant state or territory government, with levels of funding generally being determined by need. In Queensland, Victoria and South Australia, government funding was determined by a standard amount per student plus a needs-based component determined by socio-economic status scores in the time period studied. Schools in New South Wales were allocated

[8] All amounts are reported in real Australian dollars throughout this paper.

funding based on the Australian Government Education Resources Index, which allocates funding taking into account the private income of the school.[9] In the time period covered by our analysis,[10] over 80% of the funding received by public schools came from this source. The federal government supplemented this funding with a contribution equal to 10% of Average Government School Recurrent Costs (AGSRC) per student (a measure of the estimated cost of educating a student in the public system). The total amount of recurrent funding provided to each school increased each year, keeping these proportions the same. In theory, schools in the public system are completely funded by the government, but many are charging a small 'voluntary contribution' to parents to supplement this funding. This contribution, however, is not compulsory and minimal in size, averaging at only $362 per student each year.

2.3.2 Private schools

In contrast to public schools, private schools are, for the most part, funded by parent fees, charging an average of $6957 per student each year. The amount charged in the private sector, however, varies significantly from $0 to over $27,000 per student. It must therefore not be assumed that all private schools receive generous amounts of private funding. As discussed in the preceding section, private schools have also attracted government funding since the 1970s, giving parents

[9] NAPLAN scores had no formal role in school funding formulas until 2014, which is after the period studied here. See Gonski et al. (2011) for more information on the different funding mechanisms across states in the time period studied.

[10] From 2009-2012 Australian schools were funded by the 'Average Government School Recurrent Costs' system. As mentioned previously, the *Review of Funding for Schooling* was conducted in 2011 and thus our analyses should be considered alongside the findings of this report. This report argued that levels of school funding are too low and should be increased overall as well as changing to a method of allocation based on individual student need. The 'Better Schools Plan' announced following this report was due to commence 1 January 2014. Due to changes in government, the recommendations have not been fully implemented (a modified version has been put in place) and are still under consideration and a cause for debate.

greater choice in the type of education that their child receives.[11] The amount of government funding provided to students in private schools is calculated depending on the average socio-economic status of their students. On average, the amount of government funding received by private schools is about 40% of their total income, provided largely by the federal government. The state governments supplement this with a proportion of funding that is less than half of that provided by the federal government. These grants are provided to private schools to meet operating expenses in that calendar year and are largely spent on teacher salaries. A study by Williams (1985) has shown that private schools used government funding to hire extra teachers and improve teacher-student ratios, thereby driving increased enrolments in this sector.

It must be noted that private schools in this time period are taking their government funding as given, *i.e.*, exogenous, and may manage their budgets accordingly. One could argue that schools may have behaved strategically and varied their fees given their pre-existing knowledge of what government funding they would receive. This is, however, unlikely as private schools are discouraged from using government subsidies to replace parent fees and generally increase their fees each year only to reflect inflation. Furthermore, families already attending the school or prospective students have an expectation of what fees will be each year, restricting a school's ability to increase fees suddenly.

2.4 The National Assessment Program - Literacy and Numeracy

The National Education Agreement also details a system of performance reporting for schools, of which the National Assessment Program forms part. The National Assessment Program - Literacy and Numeracy (NAPLAN) was intro-

[11] See Watson and Ryan (2010).

duced in 2008 and assesses students using standardised tests in reading, writing, language conventions, and numeracy each year. At the school level, student participation in NAPLAN tests is approximately 95%, with the lowest participation rate at 93.1% for the 2011 numeracy test in South Australia. Nevertheless, no test is able to perfectly measure a student's level of achievement and thus all tests are subject to a certain amount of measurement error. However, each year the NAPLAN tests and data go through rigorous quality assurance processes in independent agencies to ensure reliability and validity of the data.

The *My School* website has also played a significant role in ensuring high data standards in the collection of financial data from schools. While some amounts published may only be estimates (due to the state and territory governments rebundling funds before passing them on to schools), the public provision of the data ensures reliability. Traditionally, transparency on government spending ensures that money cannot be misappropriated or misalloacted. All schools are able to check their entries on *My School*, and it is likely that if their budgets were not consistent with those published they would protest. State and territory governments therefore have an incentive to rebundle funding allocations in a manner that aligns with the amounts estimated by the federal government as published on *My School*. Furthermore, the calculation of income values was a significant methodological challenge faced during the implementation of this programme. Principals of independent schools argued that the data was either overstating their income or not taking their expenses into account, whereas the Australian Education Union (AEU) argued that the information understated the income of independent schools. Earlier in the development of My School 2.0, the launch had been postponed to ensure the reliability of the data, after accounting firm Deloitte believed there were methodological flaws. It is fundamental that financial data is calculated correctly, as, paired with NAPLAN results, financial data is used to assess funding requirements. It is therefore likely that the funding data produced is reliable, as each school has an incentive for their true income to be reflected on the My School website.

3 Data and descriptive statistics

3.1 The National Assessment Program - Literacy and Numeracy

We use Australian school-average data from the National Assessment Program - Literacy and Numeracy (NAPLAN) to estimate the impact of different sources of school funding on average scores achieved at the grade 3 level in reading and numeracy. Each year, all students in grades 3, 5, 7 and 9 sit tests in reading, writing, language conventions and numeracy. The school-average scores are reported by ACARA on the publicly-available *My School* website to allow for comparisons in performance between schools. The data collected for this project are unique in that information on student background and school resources are also collected, in order that comparisons only be made between similar schools. This is unlike the traditional league tables in countries such as the United Kingdom. The paradigm driving the structure of this particular test score data set, that only schools with similar resources are comparable on performance, leads us to question if resources indeed impact performance.

3.2 Sample and variable definitions

The raw data used by this study come from four years of NAPLAN testing (2009-2012) in every school in Australia, and have been provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). Given these data, we have produced a detailed panel data set comprising public and private schools and their corresponding grade 3 reading and numeracy test scores. We focus on grade 3 as this is the earliest grade with data available. Hanushek (1979) argues that production function types of empirical specifications with test scores as the outcome are most appropriate in the early grades, where reading and numeracy are the main outputs. It is arguable that later grades produce multiple outputs with a focus on a broader range of subjects and outcomes, creating problems in estimation when focusing on just reading and numeracy test scores as educational output. Results for grade 5 are produced in Appendix E for the purposes of robustness checks. Despite high levels of participation in NAPLAN, there are cases of missing data that are generally the result of testing exemptions granted to individual students due to disability or low English proficiency. While there could be differences in the number of exemptions granted between public and private schools, our results will not be significantly affected as the total fraction of exemptions granted is relatively small.

We have necessarily made further sample restrictions for the purpose of simplifying the analysis. Firstly, we focus only on four large neighbouring states: New South Wales, Victoria, Queensland, and South Australia. Due to Australia's federal system of government, school funding and governance structures differ, often greatly, across states. We analyse each state separately, thus allowing us to account for the fact that the relationship between funding and scores may vary between states. These states have been chosen as more populous states in Australia, allowing for larger sample sizes, as well as due to their differences in funding allocation and school administration policies.[12]

[12]Tasmania, the Australian Capital Territory, and the Northern Territory would have much smaller sample sizes than the states studied. While large in size, Western Australia has been omitted due to a similar policy context to New South Wales. The four remaining states all have

Secondly, as we are interested in understanding how the relationship may vary across school type, we classify schools into the following two groups: i) *Public schools*, which are any schools funded and administered by the state government, and ii) *Private schools*, which are any schools run independently from the government. Catholic schools have been omitted from our sample as they cannot be strictly defined into either of these two groups, as outlined in section 2.1. Finally, our methodology requires that we are working with a balanced panel, so we only consider schools that have recorded test scores for the entire period of study.[13] We are left with a sample of 3778 Australian primary schools, with 85% of these schools being public.

3.2.1 NAPLAN score

We analyse two separate outcome variables, the school-average NAPLAN test score in reading and the score in numeracy. NAPLAN results for each student are calculated as mean scale scores given the raw marks achieved on each test. The scale for each subject area ranges from zero to 1000 and the same scale is used from grade 3 to grade 9. The test score scale provides not only an ordinal ranking but a scale that quantifies *how much* better or worse students

differences in institutional context that may be of interest.

[13] 11.45% of public schools and 11.04% of private schools in our original data set did not have data for all four years. A small part of this is attrition due to schools opening, closing, or amalgamating. Attrition will cause bias if it unequally affects the sample in either the exposure or outcome. In our case, school openings and closures are evenly spread over school sector, states, scores achieved, and average amount of funding received and therefore does not affect our analysis. In Australia, school closures and openings are more likely to be the result of institutional factors, such as geographical redistribution, which are not analysed here. The majority of schools enter and exit the data set due to small class sizes, as scores are only reported when there are more than 5 students in the class. We already require the omission of smaller schools on the basis of economies of scale, as an analysis for schools in general will not match with the requirements of very small schools. The balancing of our panel therefore poses no threat to our sample regarding bias.

are performing from their peers.[14] This particular construction also means that scores are comparable over time. Provided the individual scores of each student, ACARA calculates the average score for each subject for each grade level in each school.

The average grade 3 scores for reading and numeracy for public and private schools in each state are presented in Table 1. It can be seen that private schools are systematically scoring higher, on average, than public schools in both reading and numeracy across all states. For example, the average reading score in New South Wales is 452 in private schools, but only 410 in public schools, a difference of 42 points. This supports the findings by Miller and Voon (2012) that private schools perform better than public schools on NAPLAN testing at the aggregate level. Further analysis shows that the differences in score between public and private schools extends to the entire distribution of scores. As illustrated in Figure 1, not only is the state average score higher in private schools, but there are also a significant number of public schools with scores well below the overall state average, illustrated by the long left hand tail. These observed distributional differences motivate our quantile regression style of analysis, which allows us to analyse how the impact of funding is changing along this distribution of scores.

The summary statistics also highlight why we need to treat these four states separately, as we can see significant differences in achievement. For example, while public school students are scoring lower on average than their private counterparts from the same state, it can be seen that Victorian public school students are achieving higher scores than private school students in both South Australia and Queensland in numeracy. Furthermore, while public school stu-

[14] As discussed in Hanushek (1979), these kind of test score scales are better suited to educational production function style analysis.

dents in New South Wales perform worse than public school students in Victoria, the averages for private schooling in these states are relatively similar. These differences in score would indicate that there may be differences in production technologies, not only between public and private schools, but also between schools in different states.

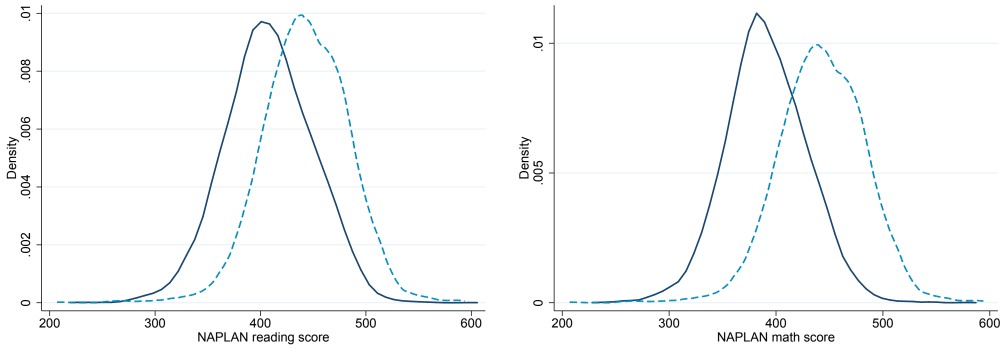
Table 1: School-average grade 3 NAPLAN scores by schooling sector

	Public	Private	Difference
Reading			
South Australia	393	427	34***
			(2.278)
New South Wales	410	452	42***
			(1.481)
Victoria	426	457	31***
			(1.717)
Queensland	385	424	39***
			(1.870)
Numeracy			
South Australia	371	399	28***
			(1.869)
New South Wales	394	429	35***
			(1.282)
Victoria	409	432	23***
			(1.573)
Queensland	371	400	29***
			(1.670)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ for t-tests for difference in means between public and private schools, calculated using OLS regression with robust standard errors to correct for heteroskedasticity. Standard error for the difference in mean in parentheses.

3.2.2 Funding

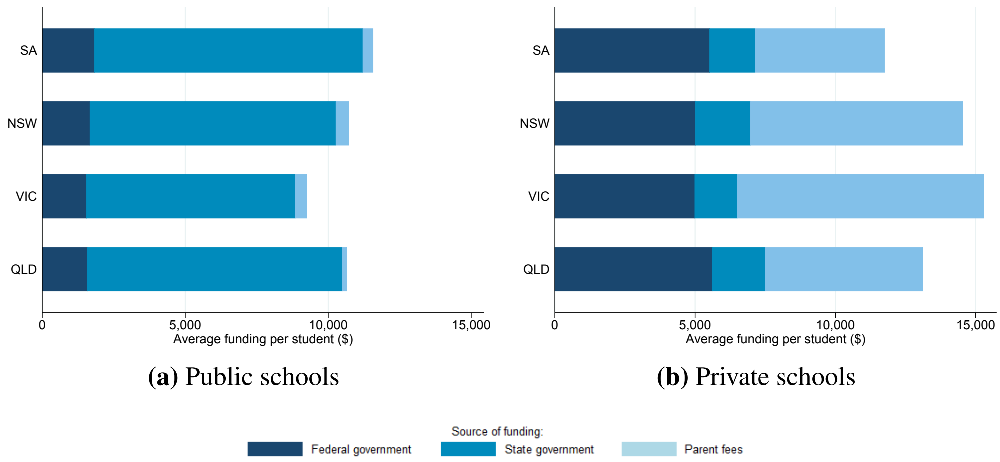
We have three key variables of interest, chosen following the three major sources of recurrent funding in schools. Australia's federal structure of governance means that all types of schools receive money from both their relevant state government, and the Australian government. In addition, schools have varying



(a) Grade 3 Reading **(b)** Grade 3 Numeracy

Figure 1: The distribution of school average NAPLAN test scores. The distributions for public schools are given by the solid navy line, and the distributions for private schools by the broken line.

abilities to charge fees to the parents of attending students. As can be seen in Figure 2, the total amount of funding per student varies, on average, between states and between public and private schools.[15] Perhaps more notably, there are significant differences in the relative amount of each type of funding that different schools receive. There are a number of issues involved with the current



(a) Public schools **(b)** Private schools

Figure 2: Relative sources of funding in Australian primary schools by state and sector. The bars above show the *average* amount of each source of funding received by schools in that particular state and sector. It must be noted that as these are averages the amounts of funding illustrated here may not be actually received by any school.

[15]Detailed summary statistics can be found in Appendix A.

process of school funding in Australia that any analysis must consider, as discussed in section 2.3. Clearly, school funding is a complex system with patterns of allocation depending on school sector, location and socio-economic background. We aim to discover how these three different sources of funding are impacting test scores in each sector and state, driven by this understanding of how funding is provided and spent. It is arguable that the current system represents significant overlap of state and federal responsibilities, so an understanding of how each type of funding is related to scores is important to determine how best to allocate money in Australian schools to improve outcomes. We therefore include variables for each type of funding. The funding information collected by ACARA is separated into four different categories: Australian government (federal) recurrent funding, state government recurrent funding, parent fees and other private funding sources. Capital funding information is also included. These last two sources of funding have been omitted from our study as they are expressed as flows rather than stocks and may be sporadically high, low or zero. We focus instead on the recurrent funding received by schools aimed at meeting operating costs, with the overall capital wealth effect in schools captured by the fixed effect.

Furthermore, we aim to find out if the differences we can see in private school test score achievement, illustrated in section 3.2.1, result from their different funding structure. It can be seen from Figure 2 that private schools receive, on average, more funding per student than public schools. This leads us to question if it is the increased resources in this school sector that are leading to improved student performance on NAPLAN, or if some other factor is driving this difference, such as self-selection. We therefore estimate the effect of each type of funding in the two sectors separately, so that we can see how production technologies may differ between the two school systems.

The preceding analysis also highlights that funding in schools is not randomly assigned. Most specifically, it is possible that the levels of parent fees charged, which vary significantly between schools, may be driven by demand for attendance at that school based on their performance in previous periods. The yearly publication of NAPLAN results on the *My School* website can be thought of as a kind of indicator of quality, which parents use when choosing a school for their child. The potential for reverse causality between funding and test scores will be accounted for in our methodology, to ensure that we are estimating the true causal effect of funding in schools.

3.2.3 Controlling for socio-economic status

We control for school background using the Index of Community Socio-Educational Advantage (ICSEA), an indicator constructed by ACARA that measures a school's average socio-educational background.[16] This measure was also used to control for socio-educational advantage by Miller and Voon (2012) who found that this indicator was the main determinant of NAPLAN score. We interpret the ICSEA as a kind of propensity score for high educational attainment. Furthermore, as discussed in the previous section, the amount of government funding received by a school is largely determined by socio-economic status. The ICSEA is therefore correlated with the funding variables and must be included to control for confounding bias. Other controls usual in the literature, such as parent education, language background, and parent employment are included in the ICSEA and thus are omitted from our specification.[17] Our estimation method condi-

[16] See www.acara.edu.au for information on how this measure is constructed.

[17] While it would be better to use the original variables, our methodology employs an MCMC process that behaves better with a smaller number of variables. A propensity score-type control

tions on the school fixed effect, thus removing the need to control for any school characteristics that do not vary over time, such as location.

4 Framework and methodology

4.1 Framework

We follow the traditional education production function approach.[18] Let

$$A_{it} = f(B_{it}, P_{it}, S_{it}, I_i), \tag{1}$$

where A_{it} measures the achievement of student i at time t, B_{it} measures student i's background, P_{it} is the effects of a student's peers, S_{it} denotes school inputs, and I_i student i's innate abilities, which do not vary over time.

We aggregate this relationship to the school-level[19] as this is the unit at which funding is allocated and spent, giving us the following reduced-form model:

$$\bar{A}_{kjt} = g(\bar{B}_{kjt}, \bar{S}_{kjt}, \alpha_{kj}), \tag{2}$$

where $\bar{A}_{kjt}$ measures the average achievement of school k in state j in year t, which is some function, $g(\cdot)$, of average student background, $\bar{B}_{kjt}$, school funding, $\bar{S}_{kjt}$, and a school fixed effect, α_{kj}, that absorbs peer effects and the school's

therefore is best in our situation.

[18]While the use of the production function approach can be disputed, as discussed in Hanushek (1979), it must be recognised that the formulation of school policy relies on systematic data collection and evaluation. These forms of study thus retain high levels of policy relevance for understanding the influence of various factors on schooling performance.

[19]Hanushek et al. (1996) have argued that well-specified linear models yield unbiased estimates for school resources when data are aggregated. In cases with the presence of omitted variable bias, aggregation may upwards bias estimates. This is kept in mind when interpreting our results. Jacob et al. (2014) have also shown that aggregate data are sufficient to address many of these kinds of school policy evaluation questions.

'innate ability' to produce good students. This may be determined by tangible measures such as levels of capital or intangible measures such as ethos and organisation. Many production function-type studies place severe restrictions on the production technology, $g(\cdot)$, namely that of *uniform input effectiveness*. By implementing quantile regression analysis, we allow this assumption to remain flexible so that the effectiveness of school inputs may vary across the score distribution, as well as between the public and private school sectors. This follows Summers and Wolfe (1977), who find that the effects of school resources are not fixed across students and schools. It must also be recognised that we cannot assume that schools are behaving efficiently in their choice of $g(\cdot)$.[20]

In using this functional form, there are a number of further assumptions that are usual in the literature.[21] Firstly, we assume *additive separability*, so that the effects of each input do not interact with each other. This is necessary so that we can include separate variables for each source of funding into our specification.[22] Many studies assume *input assignment based on fixed characteristics*, requiring that funding is allocated in each period based on fixed characteristics such as socio-economic background. Due to the formulaic funding model largely based on socio-economic status, described in section 3.2.2, we believe that this assumption is met, but we will allow for the possibility of lobbying for increased funding by schools based on test scores. The possibility of lobbying introduces a threat to identification, which we will address in section 4.4.

Most importantly, there remains a discussion in the literature regarding the

[20]Deller and Rudnicki (1993) identified production inefficiencies in elementary schools in Maine.

[21]See Harris (2010) and Todd and Wolpin (2003) for discussions on the assumptions underlying education production function models.

[22]Further work needs to be done to understand the efficiency of each type of funding at the school-budget level and how they may interact with each other, but is at present limited by data availability.

effects of previous period inputs in the current period. We assume *total decay in school input effects*. This is plausible in our case as the nature of our data means that we observe different cohorts of students each year. As recurrent funding must be spent in the year it is provided, the money spent on teacher salaries or operations in a grade 3 classroom in the preceding year will not affect students in the current year. This contrasts with studies at the individual level that follow the same students throughout a number of years.[23] In our analysis, the time dimension of the data serves instead to allow conditioning on the school fixed effect, rather than providing a value-added or cumulative-type of specification. This assumption does, however, mean that our results may not be generalisable to other grades and must be interpreted only as the contemporaneous effect of funding in a grade 3 classroom.

4.2 Econometric specification

Let A^p_{kjt} denote the average score in school k in state j at time t, where $p = 0$ if school k is public and $p = 1$ if k is private.[24] Note that we have dropped the bars to signify that these are school-average scores for ease of notation. Following equation (2), we consider the following specification:

$$A^p_{kjt} = \alpha^p_{kj} + \phi^p_{jt} + \boldsymbol{S}^{p\prime}_{kjt}\boldsymbol{\beta}_p + B^{p\prime}_{kjt}\gamma_p + u^p_{kjt}, \tag{3}$$

[23]Many of these studies assume geometric decay of varying extents to estimate either cumulative EPFs or value-added specifications. This is not necessary in our case, simplifying the assumptions that must be made for estimation considerably.

[24]To assess whether pooling public and private schools together is more appropriate, we have conducted F-tests for the significance of interaction terms with a private dummy and the funding variables in the pooled model. These are reproduced in Appendix B. The tests rejected the null hypothesis of non-significance. Therefore, the remainder of our analysis considers the estimation of separate models for public and private schools and each state. This allows for changes in production technology $g(\cdot)$, as detailed in the theoretical model.

where α_{kj}^{p} is a school fixed effect, ϕ_{jt}^{p} denotes time-specific effects that control for any shock that may have affected test scores in all schools in year t, and B_{kjt}^{p} the Index of Community Socio-Educational Advantage (ICSEA) to control for a school's proneness to achieving high (or low) test scores based on their average student background. The u_{kjt}^{p} represent idiosyncratic shocks. $\boldsymbol{S}_{kjt}^{p}$ is a vector of the three funding variables of interest: federal funding, state funding and parent fees. We are interested in identifying $\boldsymbol{\beta}_p$, to understand how funding is impacting test scores in Australian primary schools, and compare how this changes between public and private schools and across the distribution of scores.

As we are concerned with the impact of funding across the distribution of schools' test performance, we estimate quantile treatment effects (QTEs) for each of the three types of funding. Specifically, following Powell (2014, 2015, 2016), we rewrite equation (3) to produce a quantile regression estimator for panel data (QRPD) with non-additive fixed effects:

$$A_{kjt}^{p} = \boldsymbol{S}_{kjt}^{p\prime}\boldsymbol{\beta}_p(U_{kjt}^{p*}) \tag{4}$$

$$U_{kjt}^{p*} = g(\alpha_{kj}^{p}, \phi_{jt}^{p}, B_{jkt}^{p}, U_{kjt}^{p}), \quad U_{kjt}^{p} \sim \mathscr{U}(0,1) \tag{5}$$

where $\boldsymbol{S}_{kjt}^{p}$ is the vector of funding treatment variables, and $\boldsymbol{S}_{kjt}^{p\prime}\boldsymbol{\beta}_p(\cdot)$ is strictly increasing. U_{kjt}^{p*} represents proneness for the outcome (high or low achievement), an unknown function, $g(\cdot)$, of the school fixed effect α_{kj}^{p}, the time fixed effect ϕjt^{p}, and the control covariate B_{jkt}^{p}, such that the disturbance term U_{kjt}^{p} is uniformly distributed. This U_{kjt}^{p*} behaves as an aggregate term that determines rank in the distribution of the latent outcome, determined by both observed and unobserved factors. Since we observe each school in four time periods, we learn a school's proneness to high or low scores. This probability is allowed to vary across schools as well as within schools over time. Note that the QRPD method allows us to choose which variables enter the quantile function, and which are used to estimate the conditional probability. It therefore allows us to interpret the estimates in the same way as we would cross-sectional quantile estimates while conditioning on the fixed effects to aid identification.[25] That is, we are able to estimate *unconditional quantiles*, which are of greater relevance for policy analysis.[26]

We define the τ^{th} quantile of the score given the treatment $S = s$ as:

$$Q_A(\tau|\boldsymbol{S} = \boldsymbol{s}) = \boldsymbol{s}'\boldsymbol{\beta}_p(\tau), \tag{6}$$

for some $\tau \in (0,1)$. The quantile treatment effects are given by the causal effect of a change in levels of school funding, holding the chosen quantile, τ, fixed:

$$\frac{\partial Q_A(s,\tau)}{\partial s} \tag{7}$$

[25] The majority of quantile estimation methods for panel data; for example, Koenker (2004), Harding and Lamarche (2009), Lamarche (2010), Canay (2011), and Galvao Jr (2011), include an additive term for fixed effects, altering the interpretation of the estimates, *i.e.*, they must be interpreted as *conditional quantiles*. These estimators would provide estimates of the distribution of $(A^p_{kjt} - \alpha^p_{kj})|S^p_{kjt}, B^p_{kjt}$ instead of estimating the distribution of $A^p_{kjt}|S^p_{kjt}, B^p_{kjt}$. In other words, schools that are at the top of the conditional score distribution given their unobserved characteristics, may actually be at the bottom end of the unconditional score distribution.

[26] Firpo et al. (2009) provide an estimator for unconditional quantile regressions (replicated in Baltagi and Ghosh (2017)) that cannot yet incorporate instrumental variables as required by our analysis. Frolich and Melly (2013) provide an estimator for unconditional quantile treatment effects under endogeneity; however, they consider only a binary treatment variable.

Powell (2015, 2016) specifies two moment conditions for GMM estimation:[27]

$$\mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{Z}^p_{kjt}-\bar{\boldsymbol{Z}}^p_{kj})[\mathbb{1}(A^p_{kjt}\leq q(\boldsymbol{S}^p_{kjt},\tau))]\right\}=0 \qquad \text{(QRPD-GMM1)}$$

$$\mathbb{E}[\mathbb{1}(A^p_{kjt}\leq q(\boldsymbol{S}^p_{kjt},\tau))-\tau]=0 \qquad \text{(QRPD-GMM2)}$$

where $\boldsymbol{Z}^p_{kjt}$ are instruments, $\bar{\boldsymbol{Z}}^p_{kj}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}^p_{kjt}$, $\boldsymbol{S}^p_{kjt}$ are the funding treatment variables, τ is the τ-quantile of A_{kjt}, and $q(\cdot)$ is a strictly increasing function of τ. QRPD-GMM1 is a within transformation that specifies that within school variation in the instruments is used for identification.[28] QRPD-GMM2 ensures that the function $q(\cdot)$ behaves like a quantile function. The GMM estimator of $\boldsymbol{\beta}_p(\tau)$ in equation (6) solves the minimisation problem:

$$\min_{b\in\mathscr{B}} Q(b)=m(b)'W(b)m(b), \tag{8}$$

$$m(b)=\frac{1}{N}\sum_{k=1}^{N}m_k(b), \tag{9}$$

$$m_k(b)=\begin{bmatrix}\frac{1}{T}(\boldsymbol{Z}^p_{kjt}-\bar{\boldsymbol{Z}}^p_{kj}[\mathbb{1}(A^p_{kjt}\leq \boldsymbol{S}^{p\prime}_{kjt}b]\\ \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(A^p_{kjt}\leq \boldsymbol{S}^{p\prime}_{kjt}b)-\tau\end{bmatrix} \tag{10}$$

where $\mathscr{B}=\left\{b:\tau-\frac{1}{N}<\frac{1}{N}\sum_{k=1}^{N}\mathbb{1}(A^p_{kjt}\leq S^{p\prime}_{kjt}b)\leq\tau \text{ for all } t\right\}$, $b:=\boldsymbol{\beta}(\tau), W(\cdot)$ is a weighting matrix, and N is the number of schools. The parameters are restricted to $\mathscr{B}$ in order to ensure that $A^p_{kjt}\leq \boldsymbol{S}^{p\prime}_{kjt}b$ holds for approximately $100\tau\%$ of the observations in each time period. It is worth noting that the Powell (2016) method yields consistent estimates for short T and large n, as is the case in this study. This is in contrast to additive fixed effect and instrumental variable

[27] See Powell (2015, 2016) for the full assumptions and proofs for the model.

[28] Typical instrumental variable quantile regression, such as Chernozhukov and Hansen (2006) and Chernozhukov and Hansen (2008), relies on a much stronger restriction that $P(A^p_{kjt}\leq \boldsymbol{D}^{p\prime}_{kjt}\boldsymbol{\beta}_p(\tau)|\boldsymbol{Z}^p_{kjt})=\tau$.

quantile estimates from other more traditional methods. Our method for solving this optimisation problem is outlined in the following section 4.3. Issues surrounding identification are then discussed in section 4.4.

4.3 Estimation

We implement simulation-based estimation via adaptive Markov chain Monte Carlo (MCMC).[29] These methods estimate by sampling directly from the joint parameter distribution, with the aim of simulating the distribution of interest by constructing a Markov chain with the same stationary limiting distribution. This type of estimation avoids optimisation, which can pose a problem when dealing with difficult densities in cases such as ours.[30] Powell (2015) has shown that MCMC methods work well in multivariate applications of QRPD.[31]

Specifically, we implement a Metropolis-Hastings type algorithm, a variation on adaptive MCMC with vanishing adaptation[32] that uses an accept/reject rule where each draw is retained with a certain probability[33] to 'correct' an arbitrary Markov chain. If this acceptance rate of draws is too low the algorithm will get stuck in a certain region of the parameter range, whereas if it is too high the algorithm will not reach the tails of the distribution. Following Roberts et al. (1997) and Roberts and Rosenthal (2001), we initially targeted mean acceptance rates between 0.44 and 0.234. We ultimately report results from estimations with

[29] See Roberts and Rosenthal (2009) for examples, and Baker (2014) for a discussion of adaptive MCMC routines in Stata.

[30] MCMC methods have many applications to traditional statistical inference problems usually estimated by maximum likelihood or GMM procedures, see Chernozhukov and Hong (2003) for a discussion.

[31] See Stata module at Baker (2016).

[32] An example of this algorithm is reproduced in Appendix D.

[33] See Step 5 in Appendix D.

a targeted acceptance rate of 0.45[34] with 2000 draws[35] and a burn-in period of 100 for each estimation run. This relatively high acceptance rate was chosen as it is particularly difficult to traverse the support of the distribution well in our case.[36] Furthermore, as the distributions are significantly heavy-tailed, we also take logarithms of all variables with the aim of improving the mixing properties of the algorithm.[37] The point estimates are calculated by averaging the draws from this process[38] and the standard errors are derived from the variance of these draws and clustered by school.

4.4 Identification strategy

As discussed in section 3.2.2, a significant threat to identification of $\boldsymbol{\beta}_p$ rests with possible endogeneity between levels of funding and test scores achieved in previous periods. This comes from the administrative nature of our data, which means that funding has not been randomly assigned to schools but is instead determined endogenously within the system. More precisely, funding is determined by the socio-economic status of the school (the ICSEA) and by the level of funding received in the previous period. Furthermore, especially in the case of parent fees, we argue that levels of funding are determined partly by test scores in the previous period. As a result of this reverse causality between fund-

[34] The achieved acceptance rates lie between 0.414 and 0.483, with an average of 0.445.

[35] Implementing more than 2000 draws caused difficulties in deriving standard errors.

[36] As discussed in Jarner and Roberts (2007), Metropolis-type algorithms can perform poorly if local moves find it difficult to move between modes or reach significantly into the tails of the distribution. This is the reasoning behind our choice of acceptance rate, as lower acceptance rates such as 0.25 cause the algorithm to become stuck in the wrong area of the parameter space.

[37] This follows Mengersen and Tweedie (1996), who find that random-walk Metropolis algorithms for a density π are geometrically ergodic iff π has exponential or sub-exponential tails. Taking logs will ensure this condition. Taking logarithms of the funding and score variables also give useful scaling properties.

[38] As autocorrelation is a property of MCMC algorithms we retain only one in every three draws to calculate this mean.

ing and scores, the u^p_{kjt} in equation (3) may not be *i.i.d.* and uncorrelated with the regressors, *i.e.*, $\mathbb{E}[u^p_{kjt}|S^p_{kjt}, B^p_{kjt}] \neq 0$, thus potentially making the quantile estimates inconsistent.

We identify the causal effect by taking a limited-information structural modelling approach, where levels of funding in the current period are implicitly determined by funding in the previous periods through the use of instrumental variables.[39] There is little doubt that the lagged levels of funding are correlated with funding in the current period, as illustrated by Table C1 in Appendix C. It can be seen that, in most cases, lagged funding are not weak instruments.[40] The problem, however, is whether they are valid instruments. Due to data limitations regarding the dimension of time, we can only use one lag of each funding variable as an instrument. This results in a *just-identified* model, preventing us from testing the exclusion restriction to check instrument validity. However, as considered by our assumption of *total decay* in section 4.1, the money spent in a grade 3 classroom in the preceding year will have no impact upon the scores of the grade 3 students in the current year. While spending on capital is unlikely to depreciate between cohorts, recall that we focus on recurrent funding that is spent on daily operations each year. We can therefore conclude that the lagged funding variables are likely not correlated with u^p_{kjt}, and this assumption is in fact the identifying assumption of the model. To further document this identification strategy, we estimate the QRPD model with no control of endogeneity (no IVs used), and then recover the fitted residuals. See Appendix C for correlations between these predicted residuals and the demeaned lagged funding

[39]Note that the QRPD estimator avoids the usual problems involved with using lagged variables alongside a fixed effects term, as we condition on the fixed effect instead of actually estimating it.

[40]2SLS estimates using the lags as instruments are reproduced in Appendix C for interest.

variables at the 0.1 and 0.9 quantile. These highlight that the lagged levels of funding are good instrumental variable candidates across the entire distribution of the support, not just at the mean. Scatterplots shown for Victoria, as an exemplar, evidence that there is no systematic correlation between the predicted residuals and the candidate instruments. Note that we test for correlation with the demeaned lags of funding to match the first-order conditions given in Powell (2016).

5 Results

We look at the relationship between school average NAPLAN test scores and *state funding*, *federal funding*, and *parent fees*, respectively in section 5.2, section 5.3, and section 5.4. In order to keep the results concise, we only present the estimates for the 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles in the tables. We show that there is a complex and changing relationship between funding and test scores, depending on state, sector, and provenance of funding. Moreover, our results illustrate that any attempt to isolate the effect of school funding must take into account heterogeneity of impacts across the test score distribution.

5.1 A note on coefficient interpretation

The reader will notice that many coefficients are negative, which seems counter-intuitive as funding is unlikely to have a negative impact on test scores. We have interpreted these coefficients as indicating a problem with the levels of funding, which we have termed 'inefficient' referring to allocative inefficiency. We expect that the marginal effect of a dollar is non-linear across the test score distribution, and it is also true that it may be non-linear depending on the level of funding received. For example, it may be more allocatively efficient, on average, for an extra dollar in a high performing school to be given to a low-performing school instead. There is likely a point at which the marginal utility of an extra dollar is negligible, and we believe this may be driving negative coefficients. The opposite is of course also possible, that low-performing schools are receiving levels of funding that are too low to have a beneficial impact, and this could explain negative coefficients for these schools.

It will also be noted by the reader that many of the results exhibit reversals in sign and magnitude across the test score distribution. This is to be expected given the hypothesis above that the marginal effect will change over levels of funding. This is also why funding appears to have little effect on average, as the effects cancel each other out. It may be the case, for example, that the level of funding provided to a school is not the optimal amount for a given level of achievement, but is for another given level of achievement. It is likely that funding levels or policies may better target higher-achieving students than lower-achievers, or vice versa, and that this will also change depending on a state's institutional factors.

5.2 State funding

Table 2 and Table 3 present the estimates for *state funding* on reading and numeracy NAPLAN scores. It can be seen that state government funding is, for the most part, positively impacting NAPLAN scores, but with considerable heterogeneity between states and school type. This positive relationship is particularly strong in the case of Victorian public schools, where *state funding* has a positive and significant impact on scores from the 0.25 quantile and above, with a large coefficient of 1.182 on the 0.9 quantile for numeracy. Interestingly, this is a result that was missed by the simple 2SLS results in Table C6 in Appendix C, which indicate an insignificant relationship between state funding and NAPLAN scores for numeracy. Once heterogeneity in impacts across the distribution have been taken into account, however, we can see that there is indeed a strong positive relationship between *state funding* and test scores in many Victorian public schools, which a traditional mean-focused approach missed.

Conversely, we largely see a negative relationship between *state funding* and test scores in public schools in New South Wales, except at the 0.1 quantile for reading and the 0.9 quantile for numeracy. As noted in section 4.1, these negative coefficients are especially problematic given that these types of estimations at the aggregate level are most likely to have positively biased estimates. The negative coefficients would indicate that state government funding may not be working efficiently to improve basic skills in New South Wales. This finding is particularly interesting given the context that public school students in New South Wales are, on average, performing worse than those in Victoria, as we saw previously in the summary statistics in Table 1. Indeed, Victorian public school students are even performing better in numeracy than private school students in both South Australia and Queensland, despite the fact that Victoria is the state where the largest proportion of families choose private education.

The major difference between these two states in their budget allocation mechanisms is that New South Wales allocates funding and implements school policy with a traditional top-down approach,[41] whereas Victorian principals have a significant amount of power to allocate their budgets within school as they see fit. Cobb-Clark and Jha (2013) have shown that this allocative power

[41]Western Australia and the Australian Capital Territory follow the same system and have not been studied here.

improves efficiency of spending in Victorian public schools. It is therefore arguable that a top-down approach to funding is less successful than one founded in the tenets of federalism, where decision-making regarding spending is made most successfully at as local a level as possible.

Interestingly, however, the top-down approach of the New South Wales government has potentially played a role in helping the private system thrive. Our results show that private schools in New South Wales receive a significant positive benefit from their state government funding, again especially at the higher quantiles with a coefficient of 1.151 at the 0.75 and 1.216 for the 0.9 quantile for reading, and 1.666 for numeracy at the 0.9 quantile. This means that at the higher quantiles, a 1% increase of state government funding approximately translates to a 1% increase in NAPLAN score. This is not a small impact considering that these are already high-performing schools achieving NAPLAN scores at the 90th quantile. It must also be noted that this seems to be at the expense of public schools in New South Wales, which in contrast exhibit largely negative coefficients for state funding. Despite the relatively similar performance between private schools in New South Wales and Victoria, on average, we do not see a similar trend in Victorian private schools, with the coefficients on *state funding* either insignificant or negative, except at the 0.75 quantile for numeracy.

The positive impact of *state funding* on NAPLAN test scores largely continues in schools in Queensland and South Australia. In Queensland, we see some small positive effects generally around the median for both public and private schools,[42] and in South Australia we also see some positive effects. These two

[42]It must be noted that, unlike other states in Australia, Queensland lagged in developing a significant education system until the 1960s (Collins et al., 2011) and the socio-economic context of Queensland presents significant obstacles to improvement when compared to other states.

states have funding policies based on standard and needs-based components, similar to Victoria, and as a result are behaving in a similar fashion.

Table 2: Quantile Estimates for State Funding (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.261*	−0.067	0.082	−0.094	0.127***
	(0.137)	(0.054)	(0.056)	(0.138)	(0.015)
New South Wales	0.189***	−0.114***	−0.117	−0.039**	−0.335***
	(0.037)	(0.008)	(0.090)	(0.016)	(0.013)
Victoria	0.172	0.662***	0.484**	0.502***	0.553***
	(0.134)	(0.156)	(0.203)	(0.067)	(0.154)
Queensland	0.151	−0.050**	0.278***	0.251**	0.009
	(0.135)	(0.020)	(0.050)	(0.101)	(0.011)
Private schools					
South Australia	0.409	0.466***	0.951***	−0.211***	0.052***
	(0.262)	(0.056)	(0.006)	(0.029)	(0.001)
New South Wales	0.111***	0.123***	0.004	1.151***	1.216***
	(0.042)	(0.038)	(0.181)	(0.066)	(0.000)
Victoria	0.107	0.449	0.323	−0.012	−0.037***
	(0.177)	(0.546)	(0.406)	(0.010)	(0.001)
Queensland	-0.037***	0.117*	0.035***	0.175***	−0.151
	(0.006)	(0.061)	(0.003)	(0.054)	(0.272)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table 3: Quantile Estimates for State Funding (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.268	−0.095	0.819**	0.002	−0.001
	(0.244)	(0.077)	(0.405)	(0.008)	(0.003)
New South Wales	0.033	0.023	−0.041	−0.236***	0.407***
	(0.087)	(0.064)	(0.031)	(0.003)	(0.010)
Victoria	0.115	0.527**	0.384***	0.759***	1.182***
	(0.240)	(0.235)	(0.145)	(0.107)	(0.298)
Queensland	0.026	0.695***	0.658*	0.008	0.291
	(0.045)	(0.239)	(0.349)	(0.055)	(0.208)
Private schools					
South Australia	0.334***	0.017	0.947***	0.409***	0.039
	(0.048)	(0.150)	(0.002)	(0.002)	(0.026)
New South Wales	-0.081	0.469	−0.061	0.553***	1.666***
	(0.056)	(0.706)	(0.084)	(0.125)	(0.094)
Victoria	-0.214	0.066	−0.107***	0.179***	−0.281***
	(0.363)	(0.099)	(0.003)	(0.021)	(0.041)
Queensland	-0.229***	0.032***	0.612***	−0.509*	−0.772***
	(0.059)	(0.012)	(0.037)	(0.290)	(0.009)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

5.3 Federal funding

Contrary to the results for *state funding*, Table 4 and Table 5 show largely negative estimated coefficients for the impact of *federal funding* on NAPLAN test scores in both reading and numeracy. The effect of federal government funding on school-average NAPLAN test scores is unambiguous, and illustrates the limits of centralised funding allocations, casting doubt on the efficiency of federal policy in supporting schools to improve their performance.

Successive Australian governments have pledged not to decrease funding to any school, making it unlikely that an increase in test scores could be related to decreased levels of funding. Negative coefficients would therefore indicate an inefficiency in spending and illustrate the limits of the federal government to centrally allocate money to schools successfully. A possible reason for such a result could reside in the nature of the funding itself, as the spending of federal money follows strict guidelines and offers little to no flexibility regarding its allocation. Over the time period studied, federal funds were given to schools in the form of National Schools Specific Purpose Payments (SPPs),[43] which are grants made to the states bearing conditions on how they are spent. In other words, federal funding may be spent on resources such as computing facilities,[44] but this has very little to do with students' performance in NAPLAN tests (although beneficial in other ways). In contrast, state funding is spent on daily operational needs, as schools require, and can perhaps more directly impact NAPLAN test scores.

Particularly, as federal government funding is mostly allocated to private schools, we would expect to see the strongest positive relationship between *fed-*

[43] See National Commission of Audit (2014).

[44] See Gonski et al. (2011) for details on specific SPPs.

eral funding and test scores in these schools. We, however, see the counterintuitive result that there is a negative relationship between federal funding and test scores, especially in private schools in New South Wales where we can see an estimated coefficient of -1.317 at the 0.9 quantile for numeracy. This result is particularly interesting given that the estimates for state funding in these schools show a strong positive relationship. As federal government funding is a kind of ‘top-up’ funding, it is possible that state policy in New South Wales is in fact favouring private schools, and funding received from the federal government therefore has little role to play here.[45] Moreover, the only notable cases of positive coefficients for *federal funding* are at the higher quantiles for both reading and numeracy in public schools in New South Wales. As we saw previously, these schools are generally performing more poorly than public schools in Victoria and we saw negative coefficients for the estimated impact of state funding in these schools. Potentially, funding from the federal government is making up for the shortcomings of New South Wales state government funding in public schools. However, while this remains speculation, what is clear is that federal government funding generally has a minimal impact on its own, and largely acts as a ‘top up’ kind of spending that is perhaps only helpful if state government funding is not already at a high enough level.

[45] In fact, if we look at the levels of funding provided to public schools in New South Wales, they are receiving significantly more funding from the state government than Victorian schools. The median for state government funding in a Victorian public school is only \$1240 per student, compared to \$2030 per student in New South Wales.

Table 4: Quantile Estimates for Federal Funding (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.232**	0.068	−0.055***	0.032	−0.038***
	(0.090)	(0.065)	(0.020)	(0.037)	(0.015)
New South Wales	0.057***	0.056***	0.072**	0.012	0.515***
	(0.022)	(0.006)	(0.031)	(0.009)	(0.019)
Victoria	-0.272**	−0.600***	−0.405***	−0.439***	−0.402***
	(0.130)	(0.124)	(0.150)	(0.032)	(0.118)
Queensland	0.177	−0.134	−0.218***	−0.100	−0.176***
	(0.182)	(0.095)	(0.066)	(0.077)	(0.032)
Private schools					
South Australia	-0.284*	−0.634***	−0.773***	0.213***	−0.166***
	(0.151)	(0.067)	(0.003)	(0.025)	(0.043)
New South Wales	-0.092***	−0.350**	−0.544**	−1.017***	−0.655***
	(0.011)	(0.138)	(0.236)	(0.063)	(0.000)
Victoria	-0.098	−0.373	−1.545	0.013	−0.153***
	(0.105)	(0.448)	(1.001)	(0.016)	(0.018)
Queensland	0.090	−0.094***	−0.084***	−0.130***	−0.504
	(0.063)	(0.011)	(0.002)	(0.035)	(0.606)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table 5: Quantile Estimates for Federal Funding (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.313	0.052	−0.622**	0.000	0.036***
	(0.439)	(0.051)	(0.255)	(0.010)	(0.002)
New South Wales	-0.154**	−0.026	−0.012	0.129***	0.074***
	(0.077)	(0.016)	(0.029)	(0.002)	(0.007)
Victoria	-0.163	−0.416***	−0.413***	−0.589***	−1.332***
	(0.100)	(0.136)	(0.087)	(0.098)	(0.347)
Queensland	-0.277**	0.002	−0.481*	−0.003	−0.135
	(0.109)	(0.158)	(0.268)	(0.174)	(0.194)
Private schools					
South Australia	-0.360***	−0.297	−0.827***	−0.678***	−0.279***
	(0.056)	(0.215)	(0.001)	(0.004)	(0.088)
New South Wales	-0.465***	−0.391	−0.305**	−0.426***	−1.317***
	(0.124)	(0.428)	(0.154)	(0.085)	(0.069)
Victoria	0.126	−0.353**	0.032***	−0.446***	−0.110***
	(0.225)	(0.147)	(0.003)	(0.021)	(0.030)
Queensland	-0.530***	−0.129***	−0.315***	−0.369*	−1.050***
	(0.040)	(0.022)	(0.027)	(0.201)	(0.043)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

5.4 Parent fees

The results for *parent fees* are presented in Table 6 and Table 7. The main goal of looking at the results for parent fees is to find whether it is the higher parent fees in private schools that are driving their improved performance in NAPLAN testing. The results would indicate that there is generally not much difference in magnitude for the impact of parent fees between public and private schools. For example, at the 0.9 quantile, the coefficient for a private school in Queensland is 0.057 compared to 0.087 in a public school for numeracy. In fact, in some cases the coefficient for *parent fees* is in fact negative at the 0.9 quantile.[46] It is therefore not the case that extremely high levels of parent fees are always going to improve NAPLAN scores, in any type of school.

Nevertheless, there are cases where *parent fees* are having a positive impact. If we look at the two more poorly-performing states in our sample, South Australia and Queensland, these states are, for the most part, receiving a benefit from parent fees in both the public and private sectors. Overall it is clear that parent fees, although a very small fraction of the public schools' total funding, can play a key role in helping schools improve their NAPLAN test outcomes. This may be due to the school's ability to allocate funding where it is needed most, since there is no restriction as to how parent fees should be spent.[47] In some other cases, such as in private schools, the income received by parent fees may be bringing their budgets to a level similar to that of public schools, as shown in Figure 2. In that case, the majority of parent fees are likely to be spent on ordinary expenses aimed at schools' daily operations. As noted by Williams

[46]For example, both private and public schools in Victoria for both reading and numeracy.

[47]Schools are able to charge parent fees for "any specific purpose identified by the school" (see Victorian Department of Education and Training (2015)).

(1985), private schools do indeed mostly direct extra funding towards teacher salaries. For others, who charge significantly higher fees, it could just be that funding is directed towards extra-curricular activities, and not directly related with student performance in NAPLAN tests.

It is arguable that this is what is driving the successful appearance of private schools, as they have more power over their own operation and spending.[48] This conclusion is strengthened by the result we found for *state funding* in Victorian public schools, where principals have power over budget allocations. In these schools, *parent fees* do not have a significant positive impact, potentially as state government funding is already supporting their improvement.

[48]This interpretation links with the proposition of Coelli et al. (2018) that the rigid policy setting of schools may possibly be preventing school principals from changing their behaviours when faced with poor NAPLAN results.

Table 6: Quantile Estimates for Parent Fees (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.155***	0.090***	0.124***	0.096*	0.010***
	(0.057)	(0.012)	(0.043)	(0.054)	(0.001)
New South Wales	0.008	0.052***	−0.038*	0.034***	0.156***
	(0.022)	(0.009)	(0.021)	(0.010)	(0.004)
Victoria	0.036	0.019	−0.054***	−0.011	−0.228***
	(0.023)	(0.023)	(0.015)	(0.023)	(0.072)
Queensland	0.085*	0.007	0.040***	0.006	−0.042***
	(0.045)	(0.005)	(0.008)	(0.013)	(0.009)
Private schools					
South Australia	0.238*	0.138***	0.556***	0.159***	0.053***
	(0.143)	(0.015)	(0.003)	(0.017)	(0.007)
New South Wales	0.104***	0.017	0.033	0.812***	0.416***
	(0.011)	(0.023)	(0.035)	(0.034)	(0.000)
Victoria	0.168**	0.472	0.009	0.053**	−0.046***
	(0.078)	(0.502)	(0.131)	(0.021)	(0.005)
Queensland	0.030***	0.017***	0.009***	0.219***	−0.029
	(0.004)	(0.003)	(0.001)	(0.063)	(0.059)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table 7: Quantile Estimates for Parent Fees (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.034	0.110***	0.307***	0.019***	0.003***
	(0.156)	(0.020)	(0.082)	(0.003)	(0.000)
New South Wales	-0.005	0.031	0.012	0.022***	−0.151***
	(0.023)	(0.019)	(0.028)	(0.001)	(0.006)
Victoria	0.037	−0.066*	−0.072**	0.054	−0.199***
	(0.056)	(0.037)	(0.036)	(0.038)	(0.052)
Queensland	0.002	0.072*	0.257*	0.010***	0.087*
	(0.016)	(0.041)	(0.137)	(0.002)	(0.048)
Private schools					
South Australia	0.198***	−0.047*	0.503***	0.025***	0.023**
	(0.020)	(0.028)	(0.001)	(0.000)	(0.011)
New South Wales	-0.135*	0.320	0.082**	0.492***	1.516***
	(0.080)	(0.279)	(0.039)	(0.095)	(0.082)
Victoria	-0.005	−0.009	−0.009***	0.090***	−0.192***
	(0.169)	(0.058)	(0.001)	(0.018)	(0.026)
Queensland	-0.051***	0.056***	0.318***	0.181**	0.057***
	(0.006)	(0.007)	(0.007)	(0.092)	(0.002)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

6 Discussion and concluding remarks

Overall, we see that the relationship between funding and test scores varies between states, sectors, and funding types, as well as over the test score distribution, highlighting the level of complexity required when designing school funding policy in Australia. Most importantly for policy, we find that it is not always how much money a school is given, but how that money is allocated and spent within a school. Interestingly, Victorian public schools seem to be the success story of this case study, yet overall these schools receive less money on average than any other of the states and sectors in our sample. Indeed, in the context of federalism, the three sources of funding studied, *federal funding*, *state funding*, and *parent fees* line up to three levels of decision-making - national, state and school - respectively. The results of this study highlight the limits of centralised funding allocations, and in terms of policy would recommend a renewed emphasis on federalism in Australia when designing education funding policy. It is clear that there is substitutability between funding types when another source is not providing enough, meaning that money provided from one source, *i.e.*, the federal government, may in fact be preventing money from a more allocatively-efficient source, *i.e.*, the state government, being given to a school instead. This study would support the findings of unimpressive results from increased Commonwealth encroachment on education funding in the Global Financial Crisis period, and in conjunction with the needs identified by the 2011 *Review of Funding for Schooling*, make a clear case for the reform of federalism as it pertains to education policy in Australia.

Of course, it must be kept in mind that not all schools have the ability to charge high levels of parent fees, and that there are numerous other practical factors restricting school funding policy. It must also be remembered that these results hold firstly only for grade 3 classrooms, and secondly only for the outcome of NAPLAN scores, as school funding has other important functions in providing other important services to schools. However, the future focus of this policy debate needs to be on the level at which budget decisions are made. We would argue that the school-level is the best place for budget allocation to be determined, and that there should be less restrictions on how schools spend their budgets as they know best how to allocate funds to improve the performance of their own students. Further studies therefore need to develop a better understanding of how money is actually spent within schools, and how efficiently schools themselves spend their funds.

A Detailed funding summary statistics

Table A1: South Australia

	N	Mean	S.D.	Median	Min	Max
Public schools						
Federal	1248	1.82	0.63	1.70	0.96	6.93
State	1248	9.39	2.29	8.87	5.88	24.02
Parent fees	1248	0.36	0.14	0.33	0.00	1.17
Private schools						
Federal	300	5.52	1.32	5.60	2.63	9.79
State	300	1.62	0.48	1.53	0.91	3.56
Parent fees	300	4.62	4.40	2.69	0.61	17.95

Table A2: New South Wales

	N	Mean	S.D.	Median	Min	Max
Public schools						
Federal	5164	1.67	0.68	1.42	0.98	9.26
State	5164	8.60	2.79	7.89	3.21	85.64
Parent fees	5164	0.44	0.34	0.37	0.07	14.74
Private schools						
Federal	816	5.01	1.72	5.41	1.18	9.65
State	816	1.96	0.55	2.03	0.05	3.74
Parent fees	816	7.56	6.28	4.87	0.00	27.20

Table A3: Victoria

	N	Mean	S.D.	Median	Min	Max
Public schools						
Federal	3600	1.55	0.55	1.41	0.48	9.18
State	3600	7.30	1.66	6.86	4.47	23.78
Parent fees	3600	0.40	0.21	0.36	0.01	1.87
Private schools						
Federal	536	4.99	1.66	5.29	1.44	10.50
State	536	1.51	0.89	1.24	0.43	5.17
Parent fees	536	8.79	6.71	5.89	0.49	24.91

Table A4: Queensland

	N	Mean	S.D.	Median	Min	Max
Public schools						
Federal	2692	1.58	0.30	1.54	1.07	2.94
State	2692	8.91	2.80	8.23	5.41	58.44
Parent fees	2692	0.16	0.12	0.13	0.00	1.28
Private schools						
Federal	508	5.61	1.53	5.55	1.61	16.39
State	508	1.89	0.42	1.81	0.37	4.97
Parent fees	508	5.62	3.48	4.83	0.03	18.00

Note: Variables are expressed in thousands of dollars per student.

B OLS model and tests on the pooled model (public and private estimated jointly)

Table B1: Estimates for joint pooled OLS models

	(1) NAPLAN reading score	(2) NAPLAN math score
Funding		
Federal funding	-7.064***	−2.128
	(1.401)	(1.351)
State funding	-14.330***	−15.607***
	(1.861)	(1.891)
Parent fees	7.547***	6.951***
	(0.557)	(0.532)
Private schooling dummy and interactions		
=1 if Private	-15.518	−42.236
	(45.330)	(48.094)
Federal funding	-6.417*	−7.594*
	(3.642)	(4.269)
State funding	10.139***	13.283***
	(3.225)	(3.327)
Parent fees	-3.355*	−2.975
	(1.822)	(1.849)
ICSEA	0.286***	0.244***
	(0.005)	(0.005)
Observations	14864	14864
R^2	0.610	0.537

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: The logarithm has been taken of all continuous variables. Year dummies were included in each specification. Standard errors in parentheses.

C Instrument tests

Table C1: Correlation coefficients between each funding variable and its own lag

	Federal funding	State funding	Parent fees
Public schools			
SA	0.7520	0.9073	0.7762
NSW	0.6215	0.9187	0.8618
VIC	0.6114	0.8679	0.8273
QLD	0.3891	0.9331	0.5264
Private schools			
SA	0.9812	0.9134	0.9866
NSW	0.9810	0.8424	0.9908
VIC	0.9896	0.9654	0.9948
QLD	0.9792	0.7794	0.9843

Note: All coefficients are greater than 0.75 except for federal funding and parent fees in public schools in Queensland.

Table C2: 2SLS estimates for South Australia (Grade 3)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	0.028	0.144***	0.008	−0.063
	(0.043)	(0.050)	(0.051)	(0.045)
State funding	-0.035	−0.178***	−0.003	0.039
	(0.050)	(0.058)	(0.042)	(0.035)
Parent fees	0.040**	0.031**	0.017	0.007
	(0.017)	(0.016)	(0.012)	(0.010)
N	935	935	225	225
min. Eigenvalue	28.715	28.715	99.832	99.832

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C3: 2SLS estimates for South Australia (Grade 5)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	0.006	0.081***	−0.020	−0.046*
	(0.027)	(0.025)	(0.029)	(0.028)
State funding	-0.023	−0.096***	−0.081***	−0.015
	(0.030)	(0.028)	(0.024)	(0.020)
Parent fees	0.043***	0.038***	0.013**	0.018***
	(0.014)	(0.013)	(0.006)	(0.006)
N	971	971	234	234
min. Eigenvalue	36.098	36.098	74.712	94.712

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C4: 2SLS estimates for New South Wales (Grade 3)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	-0.062***	−0.015	−0.030**	−0.023
	(0.011)	(0.011)	(0.014)	(0.015)
State funding	0.001	−0.030**	−0.002	−0.025
	(0.013)	(0.012)	(0.016)	(0.018)
Parent fees	0.030***	0.029***	0.024**	0.014*
	(0.006)	(0.005)	(0.010)	(0.008)
N	3879	3879	611	611
min. Eigenvalue	259.885	259.885	196.240	196.240

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C5: 2SLS estimates for New South Wales (Grade 5)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	-0.020**	0.029***	−0.027**	−0.047***
	(0.009)	(0.009)	(0.013)	(0.014)
State funding	-0.009	−0.056***	−0.012	−0.018
	(0.010)	(0.010)	(0.016)	(0.015)
Parent fees	0.023***	0.025***	0.018**	0.001
	(0.004)	(0.004)	(0.008)	(0.007)
N	3882	3882	653	653
min. Eigenvalue	236.324	236.324	197.108	197.108

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C6: 2SLS estimates for Victoria (Grade 3)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	-0.300	0.488	0.007	0.015
	(0.301)	(0.466)	(0.019)	(0.022)
State funding	0.348	−0.630	−0.057***	−0.066***
	(0.377)	(0.584)	(0.020)	(0.021)
Parent fees	0.014**	0.003	0.003	−0.010
	(0.007)	(0.011)	(0.012)	(0.015)
N	2709	2709	402	402
min. Eigenvalue	0.945	0.945	194.779	194.779

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C7: 2SLS estimates for Victoria (Grade 5)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	-0.376	0.561	0.011	0.004
	(0.362)	(0.544)	(0.013)	(0.015)
State funding	0.458	−0.713	−0.065***	−0.051***
	(0.451)	(0.675)	(0.013)	(0.014)
Parent fees	0.013*	0.006	−0.013*	−0.016*
	(0.007)	(0.011)	(0.007)	(0.008)
N	2745	2745	411	411
min. Eigenvalue	0.809	0.809	185.417	185.417

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table C8: 2SLS estimates for Queensland (Grade 3)

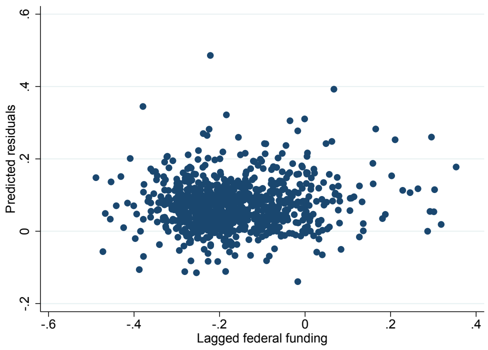
	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	0.008	−0.102**	−0.062**	−0.028
	(0.052)	(0.052)	(0.029)	(0.026)
State funding	-0.012	0.004	0.126*	0.069
	(0.013)	(0.012)	(0.073)	(0.051)
Parent fees	0.006	0.011*	−0.001	0.006
	(0.005)	(0.006)	(0.016)	(0.014)
N	1994	1994	381	381
min. Eigenvalue	39.035	39.035	73.300	73.300

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

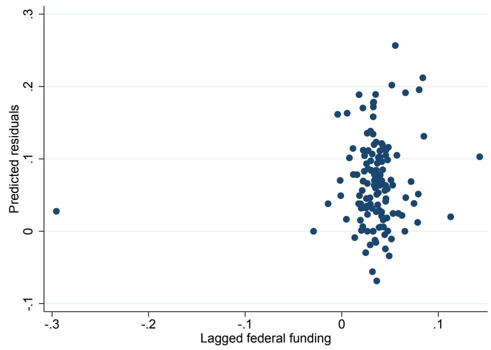
Table C9: 2SLS estimates for Queensland (Grade 5)

	Public schools		Private schools	
	Reading	Numeracy	Reading	Numeracy
Federal funding	-0.052	−0.063*	−0.035***	−0.028**
	(0.038)	(0.035)	(0.009)	(0.014)
State funding	0.005	−0.001	0.022	0.015
	(0.010)	(0.009)	(0.025)	(0.031)
Parent fees	-0.002	0.000	0.013*	0.015*
	(0.004)	(0.004)	(0.008)	(0.008)
N	2022	2022	384	384
min. Eigenvalue	34.721	34.721	256.912	256.912

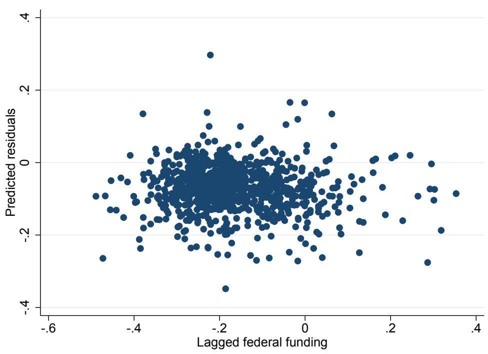
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.



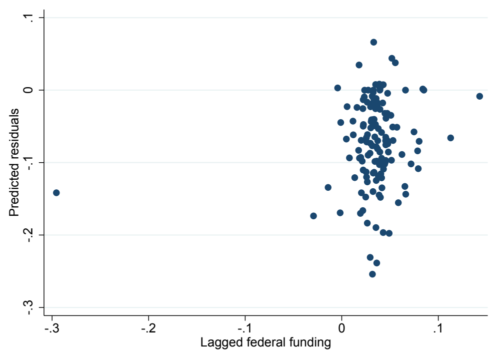
(a) Public schools at 0.1 quantile



(b) Private schools at 0.1 quantile

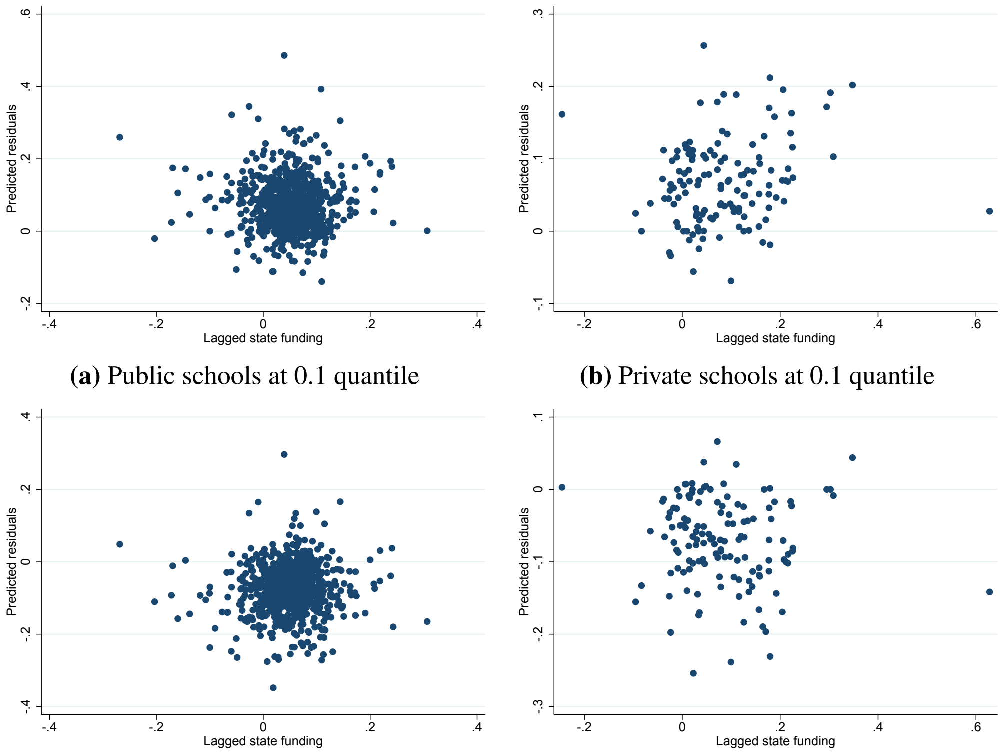


(c) Public schools at 0.9 quantile



(d) Private schools at 0.9 quantile

Figure C1: Correlation between predicted residuals and demeaned lagged federal funding. Residuals predicted for the quantile regression on year 3 reading scores. Scatter plots shown for Victoria in the year 2012.



(a) Public schools at 0.1 quantile

(b) Private schools at 0.1 quantile

(c) Public schools at 0.9 quantile

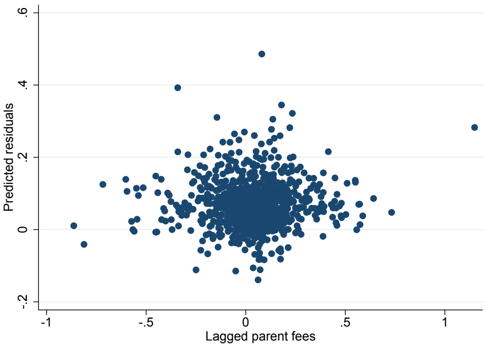
(d) Private schools at 0.9 quantile

Figure C2: Correlation between predicted residuals and demeaned lagged state funding. Residuals predicted for the quantile regression on year 3 reading scores. Scatter plots shown for Victoria in the year 2012.

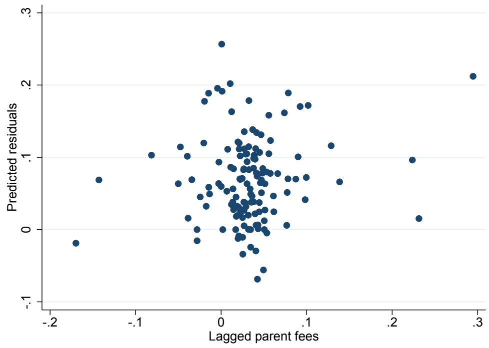
Table C10: Correlation coefficients between predicted residuals and the demeaned lagged funding variables for year 2012

	0.1 quantile			0.9 quantile		
	Federal	State	Parent fees	Federal	State	Parent fees
Public schools						
SA	0.0928	0.0313	0.0594	0.0391	0.0621	−0.0994
NSW	−0.0524	0.0289	0.0347	0.0188	−0.0203	0.0058
VIC	0.0900	−0.0256	0.0350	−0.0641	0.0426	0.1423
QLD	−0.0555	−0.0432	0.0018	0.0297	0.0386	−0.0631
Private schools						
SA	−0.1360	−0.0004	0.1561	−0.0190	0.0524	0.1584
NSW	0.0749	−0.0365	−0.0666	0.0061	−0.2056	−0.1346
VIC	0.0975	0.1899	0.1247	0.1548	−0.0837	0.0550
QLD	0.0037	−0.2450	−0.3094	0.0460	−0.2839	0.1568

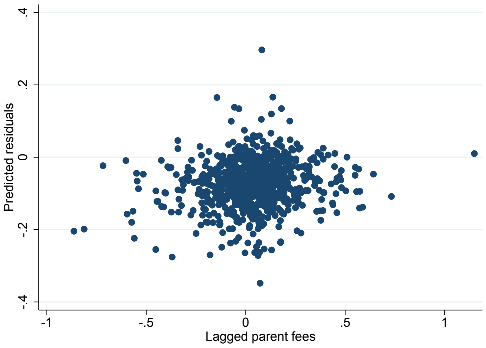
Note: Residuals predicted for the quantile regression on year 3 reading scores.



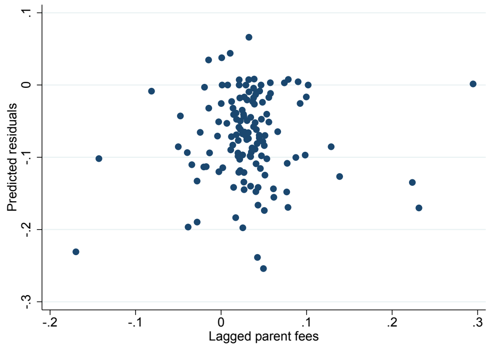
(a) Public schools at 0.1 quantile



(b) Private schools at 0.1 quantile



(c) Public schools at 0.9 quantile



(d) Private schools at 0.9 quantile

Figure C3: Correlation between predicted residuals and demeaned lagged parent fees. Residuals predicted for the quantile regression on year 3 reading scores. Scatter plots shown for Victoria in the year 2012.

Table C11: Correlation coefficients between predicted residuals and the demeaned lagged funding variables for year 2011

	0.1 quantile			0.9 quantile		
	Federal	State	Parent fees	Federal	State	Parent fees
Public schools						
SA	−0.0178	0.0058	0.1944	0.1385	0.0287	−0.1125
NSW	0.0597	−0.0907	0.0425	0.0736	−0.0533	0.0373
VIC	−0.0556	0.0522	−0.0220	0.0816	0.0479	0.0110
QLD	−0.0185	0.0993	0.1207	0.0209	−0.0286	−0.0030
Private schools						
SA	−0.3029	0.2484	−0.0663	−0.3525	0.2388	−0.1138
NSW	0.0923	0.0873	0.0001	0.1255	0.0745	−0.0039
VIC	−0.0276	−0.1486	0.1193	0.0072	0.1813	0.1098
QLD	−0.3424	−0.0165	−0.1602	0.1798	−0.1026	−0.4732

Note: Residuals predicted for the quantile regression on year 3 reading scores.

Table C12: Correlation coefficients between predicted residuals and the demeaned lagged funding variables for year 2010

	0.1 quantile			0.9 quantile		
	Federal	State	Parent fees	Federal	State	Parent fees
Public schools						
SA	0.0831	0.0727	0.1536	0.2786	0.1121	0.0353
NSW	−0.1687	0.0521	−0.0766	0.0750	−0.0522	−0.0342
VIC	−0.0409	−0.0059	−0.0530	0.0884	−0.0010	−0.0425
QLD	0.0332	0.0650	−0.0327	−0.0580	0.0998	−0.0308
Private schools						
SA	0.1501	−0.1498	−0.1468	0.0281	0.0514	−0.0887
NSW	−0.0303	0.0187	−0.0609	−0.0221	0.0294	−0.0371
VIC	0.0522	−0.1325	−0.1257	−0.0193	0.0086	−0.0901
QLD	0.0180	0.2047	0.3192	−0.1113	0.3080	0.0011

Note: Residuals predicted for the quantile regression on year 3 reading scores.

D Adaptive MCMC with vanishing adaptation

Table D1: Overview of an adaptive Metropolis-Hastings algorithm with a multivariate normal proposal and vanishing adaptation

Adaptive MCMC algorithm with normal proposal and vanishing adaptation
1: Set starting values $X_0, \mu_0, \Sigma_0, \lambda_0, \alpha^*, \delta(\delta > 0)$, and draws T.
2: Set $t = 0$ and repeat steps 3-10 while $t \leq T$:
3: Draw a candidate $Y_t \sim MVN(X_t, \lambda_t \Sigma_t)$.
4: Compute $\alpha(Y_t, X_t) = \min\left[\frac{\pi(Y_t)}{\pi(X_t)}, 1\right]$
5: Set $X_{t+1} = Y_t$ with probability $\alpha(Y_t, X_t)$,
$X_{t+1} = X_t$ with probability $1 - \alpha(Y_t, X_t)$.
6: Compute weighting parameter $\gamma_t = \frac{1}{1+t^{\delta}}$.
7: Update $\lambda_{t+1} = \exp[\gamma_t(\alpha(Y_t, X_t) - \alpha^*)]\lambda_t$.
8: Update $\mu_{t+1} = \mu_t + \gamma_t(X_{t+1} - \mu_t)$
9: Update $\Sigma_{t+1} = \Sigma_t + \gamma_t[(X_{t+1} - \mu_t)(X_{t+1} - \mu_t)' - \Sigma_t]$
10: Increment t.
Output: the sequence $\{X_t\}_{t=1}^{T}$

Source: Reproduced from Baker (2014).
Note: In step one, the algorithm starts with initial value, X_0, an initial variance-covariance matrix for proposals, Σ_0, an initial value of the scaling parameter, λ_0, and a targeted acceptance rate α^*. δ is an averaging or damping parameter that controls how quickly the impact of the tuning mechanism decays through the parameter γ_t.

E Grade 5 Results

The results for grade 5 are reproduced in this Appendix for the purposes of robustness checks. While the magnitudes differ, we can see the same general patterns as in the grade 3 results. For example, the estimates for *federal funding* remain negative, except for some positive estimates for reading in public schools in New South Wales. The positive estimates for numeracy in South Australian public schools are also reproduced by the results for grade 5. The patterns for *parent fees* also match those exhibited by the grade 3 results, with significant positive effects save for some negative estimate at the higher quantiles in Victorian public schools for reading and numeracy. Furthermore, we see negative coefficients at the 0.9 quantile for numeracy in public schools in New South Wales for both grade levels. There are, however, some changes in sign at the lower quantiles for numeracy regarding *parent fees*.

Our key results for *state funding* are also largely supported by the results for grade 5. Firstly, private schools in New South Wales are still exhibiting strong positive coefficients for *state funding*, indicating that conditions in this state seem to conducive to private schools succeeding. We also identify again a strong positive effect for *state funding* in Victorian public schools. The positive effects for Queensland, however, are less clear in the grade 5 results than those for grade 3.

It must be remembered that the education production function theory behind our study relies on reading and numeracy being the major outputs of education. As noted by Hanushek (1979), at the older grade levels the emphasis becomes less on basic skills and more towards varied outputs in a manner of different subjects. For this reason, it is unsurprising that we see slight differences between the results for grade 3 and grade 5. This information, however, is still of interest to the policy-maker in this domain. It must therefore be noted again that the main results of our paper are the contemporaneous effect of funding *in a grade 3 classroom*, and more work must be done to understand the role of funding in older classrooms with more varied learning outputs.

Table E1: Quantile Estimates for Federal Funding (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.099	−0.401***	−0.048	0.031***	0.057*
	(0.719)	(0.075)	(0.231)	(0.006)	(0.033)
New South Wales	-0.033***	−0.003	0.019***	0.446***	0.041
	(0.009)	(0.006)	(0.006)	(0.049)	(0.099)
Victoria	-0.307***	−0.018	−0.167***	−0.277**	−0.216***
	(0.099)	(0.177)	(0.019)	(0.114)	(0.022)
Queensland	0.067***	−0.094***	0.008	0.133	−0.848***
	(0.012)	(0.020)	(0.116)	(0.355)	(0.322)
Private schools					
South Australia	-0.050	−0.386	−0.056***	−0.520***	0.055
	(0.157)	(0.399)	(0.006)	(0.000)	(0.060)
New South Wales	0.146	−0.239	−0.384**	−3.161***	−2.040**
	(0.298)	(0.361)	(0.175)	(0.510)	(0.839)
Victoria	-0.139*	−0.308***	0.303	−0.142***	−0.012***
	(0.078)	(0.072)	(0.349)	(0.003)	(0.002)
Queensland	-0.558	0.837	−0.562	−4.052	0.003
	(1.386)	(0.757)	(0.727)	(3.024)	(0.002)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table E2: Quantile Estimates for Federal Funding (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.239	−0.171	−0.465**	0.082***	0.330***
	(0.369)	(0.111)	(0.213)	(0.013)	(0.036)
New South Wales	-0.017	0.016	−0.001	1.232**	0.538***
	(0.018)	(0.012)	(0.006)	(0.554)	(0.060)
Victoria	-0.244*	−0.666	−0.670	−1.226	−0.756***
	(0.138)	(0.435)	(0.791)	(1.054)	(0.102)
Queensland	-0.196**	−0.026***	−0.206***	−0.764***	−0.834***
	(0.078)	(0.006)	(0.013)	(0.194)	(0.182)
Private schools					
South Australia	-0.192***	0.104	−1.439***	−1.116***	−0.363***
	(0.021)	(0.895)	(0.444)	(0.023)	(0.115)
New South Wales	-0.277***	−0.037	−1.111***	−1.647***	−0.287***
	(0.075)	(0.773)	(0.267)	(0.450)	(0.105)
Victoria	-0.107**	−0.063	−0.281**	−1.317***	−0.011
	(0.054)	(0.128)	(0.140)	(0.000)	(0.010)
Queensland	-0.094***	0.006	−1.480***	−0.333	−0.171***
	(0.025)	(0.011)	(0.550)	(0.322)	(0.050)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table E3: Quantile Estimates for State Funding (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.151	0.384***	−0.080	−0.004	0.000
	(0.606)	(0.105)	(0.128)	(0.004)	(0.031)
New South Wales	-0.049***	−0.037	−0.067***	0.019*	0.024***
	(0.014)	(0.039)	(0.013)	(0.010)	(0.006)
Victoria	0.271**	−0.028	0.172***	0.283***	0.204***
	(0.135)	(0.196)	(0.021)	(0.071)	(0.008)
Queensland	-0.048***	0.024	−0.041	−0.200	−0.170***
	(0.015)	(0.044)	(0.066)	(0.538)	(0.053)
Private schools					
South Australia	-0.130	−0.522	−0.067***	0.550***	−0.001
	(0.088)	(0.896)	(0.002)	(0.000)	(0.036)
New South Wales	-0.112	−0.209	0.033	2.306***	1.570**
	(0.270)	(1.203)	(0.083)	(0.381)	(0.624)
Victoria	0.214**	0.204***	0.370	0.168***	−0.032***
	(0.109)	(0.075)	(0.355)	(0.003)	(0.001)
Queensland	0.270	0.155	0.382	1.296	0.000
	(0.693)	(0.711)	(0.358)	(1.050)	(0.008)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table E4: Quantile Estimates for State Funding (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	-0.073	0.285*	−0.142	0.039	−0.311***
	(0.618)	(0.165)	(0.333)	(0.026)	(0.037)
New South Wales	-0.092***	−0.062**	−0.015	−0.316**	0.053***
	(0.028)	(0.028)	(0.037)	(0.129)	(0.008)
Victoria	0.248	0.965	0.716	1.475	1.067***
	(0.277)	(0.591)	(0.773)	(1.291)	(0.119)
Queensland	0.092	−0.035***	0.047*	0.311***	1.138***
	(0.073)	(0.002)	(0.027)	(0.089)	(0.248)
Private schools					
South Australia	-0.016***	−0.568	0.258***	−1.961***	0.090***
	(0.002)	(0.641)	(0.082)	(0.042)	(0.027)
New South Wales	0.063***	0.344	0.843***	2.849***	0.238*
	(0.022)	(1.076)	(0.202)	(0.817)	(0.125)
Victoria	0.071	0.198	−0.098	0.415***	0.008
	(0.061)	(0.259)	(0.114)	(0.000)	(0.008)
Queensland	0.109**	0.028	−0.108	−0.591	0.112**
	(0.048)	(0.022)	(0.114)	(0.473)	(0.045)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table E5: Quantile Estimates for Parent Fees (Reading)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.049	0.138***	0.083**	0.018***	0.025*
	(0.110)	(0.026)	(0.040)	(0.002)	(0.015)
New South Wales	0.038***	0.019	0.027***	0.027***	0.015
	(0.009)	(0.012)	(0.001)	(0.001)	(0.018)
Victoria	-0.037	0.001	−0.012***	−0.014	−0.200***
	(0.030)	(0.017)	(0.003)	(0.017)	(0.009)
Queensland	0.005*	0.019***	0.036	0.068	−0.045***
	(0.003)	(0.005)	(0.033)	(0.121)	(0.013)
Private schools					
South Australia	0.032***	0.057	0.008***	0.282***	0.063**
	(0.002)	(0.042)	(0.001)	(0.000)	(0.029)
New South Wales	0.130**	0.004	0.021	1.337***	0.671***
	(0.055)	(0.507)	(0.029)	(0.213)	(0.204)
Victoria	0.128***	0.220***	−0.014	0.235***	0.004***
	(0.042)	(0.064)	(0.151)	(0.005)	(0.001)
Queensland	0.032	0.117	0.410	−0.408	0.019***
	(0.043)	(0.145)	(0.362)	(0.312)	(0.001)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Table E6: Quantile Estimates for Parent Fees (Numeracy)

	Quantile estimates				
	0.10	0.25	0.50	0.75	0.90
Public schools					
South Australia	0.022	0.178***	0.215	−0.004	−0.162***
	(0.173)	(0.025)	(0.152)	(0.009)	(0.021)
New South Wales	0.025***	0.007	0.026***	−0.028	−0.200***
	(0.004)	(0.019)	(0.004)	(0.029)	(0.027)
Victoria	-0.030	0.037***	−0.130	−0.153	−0.192***
	(0.045)	(0.013)	(0.124)	(0.287)	(0.029)
Queensland	0.066***	0.004***	0.012***	0.079**	0.421***
	(0.021)	(0.002)	(0.004)	(0.031)	(0.093)
Private schools					
South Australia	0.031***	−0.195	0.018***	0.257***	0.057***
	(0.002)	(0.438)	(0.006)	(0.005)	(0.006)
New South Wales	0.123***	0.289	0.254***	1.376***	0.207*
	(0.026)	(0.503)	(0.055)	(0.380)	(0.107)
Victoria	0.042*	−0.042	−0.009	−0.346***	0.031***
	(0.021)	(0.062)	(0.061)	(0.000)	(0.005)
Queensland	0.086***	0.033***	0.315***	0.173	0.088***
	(0.006)	(0.001)	(0.059)	(0.143)	(0.027)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses.

Statement of Authorship

Social disadvantage in Australian public and private schools: A control function approach
Submitted for publication (under revision).

Author

Name: Sarah Cornell-Farrow
Overall percentage (%): 100%
Certification: This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

25 January 2019

Social disadvantage in Australian public and private schools: A control function approach*

Sarah Cornell-Farrow†

The University of Adelaide, School of Economics, Adelaide, SA, Australia, 5005.

Abstract

This study examines the determinants of low educational achievement in Australian schoolchildren. Using a control function approach that addresses issues related to self-selection on unobservables into the private schooling sector, we estimate the impact of student background characteristics on the likelihood of meeting the national minimum standard on NAPLAN testing in reading and numeracy. Next, we analyse how private schooling impacts this likelihood for students from various socio-educational backgrounds. We find that students with a language other than English or an Indigenous background are more likely to achieve NAPLAN scores below minimum standards, as well as students with a parent who did not complete year 12, does not have a university degree or is not employed. Moreover, private schooling may make some kinds of students more likely to perform above the minimum national standard on NAPLAN.

Keywords: National testing, control function, ordered choice models, education, Australia.
JEL Classification: C31, C34, I24.

1 Introduction

Education is often seen as providing individuals with the skills to lead fulfilling lives, and countries with the tools for a productive economy.[1] Nevertheless, a

*This research is supported by an Australian Government Research Training Program (RTP) Scholarship. Thanks must also be given to the Australian Curriculum, Assessment and Reporting Authority for the provision of the data utilised by this study. Thank you to Firmin Doko Tchatoka, Gareth Myles, Robert Garrard, Julie Moschion, participants at the 2018 Australian Conference of Economists, participants at the 31st PhD Conference in Economics and Business, and four anonymous referees for useful comments that have improved this paper. Stata codes for this paper may be found at Github.com/SarahCornell-Farrow

†Email: sarah.cornell-farrow@adelaide.edu.au

[1] See Barro (1991); Hanushek and Woessman (2008); Hanushek (2013).

significant proportion of children across the globe are not meeting basic literacy and numeracy standards.[2] While it is well-known that individual-level factors such as household income, gender, ethnicity, and educational achievement of parents play a large role in determining outcomes,[3] the role that school-level factors play in student educational achievement are less clear.[4] First, as discussed in Hanushek (2003), the role that school resources play in educational outcomes remains controversial. Second, the role that public or private provision of schooling plays in educational achievement is also contested. In the United States, Elder and Jepsen (2014) found that higher outcomes in Catholic schools are solely due to selection bias, a finding supported by Gibbons and Silva (2011) in the United Kingdom. Chudgar and Quin (2012) also conclude that private schooling may not have an unequivocal positive effect. In contrast, Lefebvre et al. (2011) found that switching from a public to a private high school significantly improves student test scores in Canada. Furthermore, Altonji et al. (2000) find that attendance at Catholic high schools increases the probability of graduating high school; however, they find little evidence for test scores. Green et al. (2011) found that the observable characteristics of children attending private schools has not changed over time in private schools, yet educational achievement has increased from these schools, indicating that it may not be social or economic advantage driving improved outcomes in private schools.

This study investigates the determinants of poor educational achievement in Australian children, as well as analysing if private schooling may shield students

[2]For an overview of testing practices and achievement in OECD countries see Morris (2011).

[3]See Bedard and Cho (2010); Alderman et al. (1996); Behrman and Knowles (1999); Card and Payne (2002); Cobb-Clark and Nguyen (2012); Dickerson et al. (2015); Fryer and Levitt (2010); Gibbons et al. (2013); Fryer and Levitt (2004); Jensen and Rasmussen (2011); Nicoletti and Rabe (2013); Shafiq (2013); Stevens and Schaller (2011); Suziedelyte and Zhu (2015).

[4]See, for example, Ellison and Swanson (2016).

with disadvantaged characteristics from poor performance. Australia makes a particularly interesting setting as educational quality is relatively high, yet socially disadvantaged students quickly fall far behind their peers with similar capabilities (Goss and Sonnemann, 2016). In Australia, the literature on the differences in performance between public and private schools remains limited and inconclusive. Furthermore, few papers have focused on the National Assessment Program - Literacy and Numeracy (NAPLAN) set of data, which assesses students through primary and high school, a crucial time period where policy may still be effectively targeted at students for their improvement. Nghiem et al. (2015) used data from the Longitudinal Study of Australia's Children (LSAC) to estimate the gain from attending private school on a child's test scores, as well as their non-cognitive outcomes such as social skills. They found that private schools do not produce better outcomes when compared to public schools. Contrary to this result, Miller and Voon (2012) found that private schools scored higher on standardised tests than public schools. Nonetheless, they also concluded that these better results are the result of selection bias, corroborating the findings of Elder and Jepsen (2014) in the United States and Gibbons and Silva (2011) in the United Kingdom.

It is well-accepted that student background determines test scores, with Miller and Voon (2011) finding these measures to be strong determinants of NAPLAN scores. Furthermore, Cobb-Clark and Nguyen (2012) found differing levels of educational advantage between Australian-born, English-speaking migrant, and non-English-speaking migrant students. There is, however, little understanding as to whether different types of schooling can shield students from the effects of disadvantage, or if the two main schooling sectors in Australia simply differ in their levels of disadvantage in the first place. Given that low socio-economic status is linked to poor educational performance, it is necessary to understand policies that can be aimed at these 'at risk' children to improve their educational outcomes.

In order to do this, first we estimate the impact of student background characteristics on the likelihood of meeting the national minimum standard in reading and numeracy using an individual-level NAPLAN data set. Second, we analyse how private schooling may change the probability that a student with certain disadvantaged characteristics will meet minimum standards. This paper focuses on low-performing and socio-economically disadvantaged students and aims to understand if choice in schooling may be able to improve their outcomes, rather than focusing on overall improvement.

Given that students or parents tend to self-select into private schooling based on unobservable characteristics such as ability, we propose the control function approach of Wooldridge (2015) to achieve identification. Estimation takes a two-stage approach, which provides a simple way to estimate our fairly complicated model with a binary endogenous variable and a multinomial ordered outcome. Adding a control function, estimated in a first stage, renders the endogenous explanatory variable exogenous in the second-stage equation. Implementing the control function approach should therefore ensure identification of the effect of private schooling. A more traditional way to address sample selection into private schooling would be to implement a Heckman (1979) correction.[5] However, as we have non-linear models in the first and second stages due to limited dependent variables,[6] this approach would cause the estimates

[5] Terza et al. (2008) discusses the differences between two-stage predictor substitution such as the Heckman correction and two-stage residual inclusion such as our method in the context of the health economics literature. Other methods include work based on Lee et al. (1980) and De Luca and Perotti (2011).

[6] Semykina and Wooldridge (n.d.) have developed a semi-parametric estimator for panel

to be inconsistent. Instead, the control function approach allows us to identify the parameters under similar kinds of conditions to instrumental variable estimation. In fact, for linear endogenous explanatory variables the control function approach yields the 2SLS estimator. However, for non-linear models the control function approach offers significant efficiency advantages over standard IV methods, and delivers more reliable estimates of the partial effects (although the assumptions for consistency of the partial effects are somewhat nonstandard and controversial) (Wooldridge, 2015). Moreover, control function methods are computationally simpler and require fewer assumptions than full maximum likelihood procedures.[7]

This paper makes several contributions to the literature on public and private schooling. Importantly, we focus on *'at risk'* students, that is, students that are already or who are in danger of not meeting the Australian minimum educational standard. The availability of a large individual-level data set allows us to directly control for background characteristics in a precise way. We can link parent data directly to individual students instead of working at a school-average level, thus accounting for heterogeneity within schools instead of simply across schools. Furthermore, as the data represents close to a census of all Australian children in the calendar year 2013, the large sample size allow us to very precisely estimate small impacts of student background characteristics on outcomes. This paper provides, to our knowledge, the first systematic Australia-wide study of the determinants of poor educational achievement in schoolchildren across both the public and private schooling sectors.

data.

[7]For examples of the implementation of control function approaches in the applied literature, see Liu et al. (2010); Castells-Quintana and Royuela (2017); Di Porto et al. (2017); Semykina (2017); Yeung (2017); Bourlass et al. (2018); Le Van et al. (2018).

Ultimately, we find that students with a language background other than English or an Indigenous background are at risk of poor performance in NAPLAN. Moreover, parent background factors have an impact on student achievement, as students with a parent who did not complete year 12 or university studies, or is not employed, are more likely to achieve below minimum standards. Given the results of the study of predicted probabilities, we find that private schooling makes certain students more likely to surpass achievement standards, but not others. Specifically, students at the margin of poor achievement may most benefit from private schooling. This result has significant ramifications for the literature on school choice, indicating that matching students to sectors may play a role in their achievement capabilities. The impact of disadvantaged characteristics on student achievement need not be as great as we often see, and more can be done to enable these students to succeed. Sensitivity analyses on various specifications and subsets of the data show the robustness of these results, and can be found in the Appendix.

The paper proceeds as follows. Section 2 gives a discussion of the data, followed by the theoretical model in Section 3. Section 4 outlines our estimation strategy. Section 5 presents the results and Section 6 presents a study of predicted probabilities. Section 7 concludes.

2 Data and descriptive statistics

This paper uses individual-level data from the National Assessment Program - Literacy and Numeracy (NAPLAN)[8] to estimate the role private schooling plays in enabling students to meet minimum educational standards in Australia. The

[8]Thanks must go to the Australian Curriculum, Assessment and Reporting Authority for the provision of these data.

NAPLAN data is unique in providing standardised test results to measure educational outcomes for the universe of students in Australia. Each year, students in grades 3, 5, 7, and 9 sit tests in reading, writing, language conventions, and numeracy. The results for each student are reported to schools and parents on a scale that shows how that student has performed in the tests in comparison to established national standards. Alongside these test scores, data is collected by the Australian Curriculum Assessment and Reporting Authority (ACARA) on parent background, language background other than English status and other measures.

Specifically, we use individual-level student data from the testing years 2013 and 2014. As tests are sat by individual students every two years, no students will appear in the data set twice so we have pooled these observations giving an initial sample of 2,235,804 students in 9249 schools. Necessarily, we have made a few sample restrictions, generally due to missing data. Missing data is usually the result of non-reporting by parents on school enrolment forms or testing documents. The omission of these students may therefore pose a problem if non-respondents share similar characteristics. We also keep only students that have been in the same school for at least two years to ensure that they have not recently swapped between the public and private systems. We are left with an actual sample of 471,453 students in 8160 schools, with 45% of the students in the sample attending private school.[9] In the original sample, 35% of students

[9]Much of this significant decrease in observations is due to omitting those without a score in the previous period. The reasoning behind this is that we do not have enough information about these students to know if they have been in the public or private sector for a long enough period of time for the schooling to have had an effect. For example, many students in Year 7 will not have a previous score as they changed schools between Year 5 and Year 7 when going to high school. All students in grade 3 are also omitted for this reason. This is why we can observe significant drops in the number of students, but not in the number of schools. While dropping these observations may introduce bias if they all share similar characteristics (for example, if all students swapped from public school to private school between year 5 and year 7), this is

attended private school so the actual sample may not be representative of the true population regarding the schooling type variable. This would reflect the fact that it is harder to collect parent background data from public school students than from private school students, where these types of information are a necessity upon enrolment. Given this, we have accordingly weighted our samples by the inverse of the probability of being sampled.[10] This will allow valid inference about the population, not just the sample.

The outcome variable *achievement*, a_i, is an ordered limited dependent variable, where each individual student is classified into one of three groups based on their raw NAPLAN score: 1) not meeting minimum standards, 2) just meeting minimum standards, and 3) above minimum standards. We have classified students into these groups given the raw cutoff scores for each band in each grade in each year, using the score equivalence tables provided by ACARA.[11] We analyse student achievement in both the reading test and the numeracy test,[12] with the analysis undertaken separately for each domain due to differences in achievement between these subject areas. As these bands are a sliding scale constructed to be comparable over time for an individual student, this particular construction of the achievement variable means that we do not need to further account for a student's grade level in the specification. By limiting the dependent variable to these three groups, we increase the policy relevance of our res-

unlikely. Minimising bias caused by students not being in the school for long enough to be considered a representation of that sector is more crucial. In Appendix H we take a randomised sample of the data to ensure that the results are not being biased by sample definition.

[10]This was achieved using *pweight* in Stata to ensure means and variances are not affected due to the non-random sampling. The sample is weighted by the inverse of the probability that they are sampled as this number reflects how many individuals they represent in the population.

[11]See Appendix A.

[12]The NAPLAN reading test involves a reading comprehension test on a range of text types, designed to test a student's understanding of English in context. The numeracy test is made up of a multiple-choice section and a written response section, focusing on the areas of number and algebra, measurement and geometry, and statistics and probability.

ults as we are able to target students that are not meeting minimum standards or at risk. Secondly, as NAPLAN is often seen as an imperfect measure of actual underlying student achievement, this methodology removes systematic bias potentially introduced into the model by measurement error in raw NAPLAN scores. While NAPLAN may not perfectly measure a student's actual achievement levels, the chosen bands represent the illustration of particular skills as set out in the National Curriculum. Students must therefore demonstrate or not demonstrate these specific skills in order to be located within these bands, meaning that we can best model this outcome discretely. The proportion of students

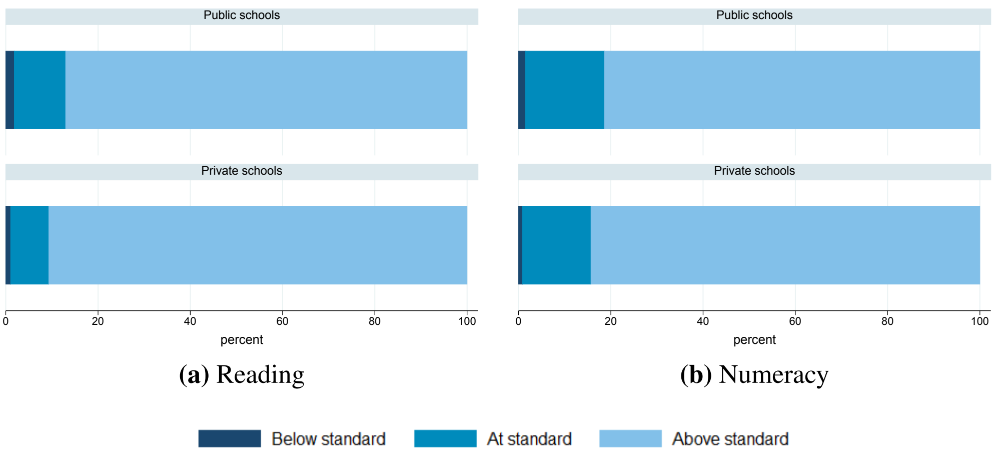


Figure 1: Student achievement by school sector

in each achievement group are illustrated by Figure 1, and compared between public and private schools. It can be seen that in private schools a higher proportion of students are achieving above standards, with 91% of students achieving in this band for reading compared to only 87% in public schools. As we shift to the lower achieving bands, the share of private school students falls, with 1.05% of private school students in the 'at minimum standards' group and 1.84% of public school students. These broad patterns are mirrored by the results in numeracy. The figures are reported in Table C1 in Appendix C, and a simple *t*-test procedure shows that, within each achievement group, the proportions of public and private schoolchildren are significantly different from each other at the 1% level. These stylised facts would lead one to believe that students in private schools are less likely to be scoring in the lower achievement bands than those in public schools, providing evidence that the findings of Miller and Voon (2012) at the school-average level are also supported at the individual student level.

Table 1: Student background characteristics by schooling sector

	Public	**Private**
Parental characteristics		
Mother completed year 12	0.6315	0.7305***
Mother holds a bachelor's degree	0.2769	0.3787***
Mother employed	0.7106	0.8253***
Father completed year 12	0.5428	0.6657***
Father holds a bachelor's degree	0.2466	0.3573***
Father employed	0.9160	0.9627***
Student characteristics		
Indigenous	0.0273	0.0104***
LBOTE	0.2379	0.2081***
Female	0.4817	0.4943***
Attends school in city	0.7428	0.7976***
Attends school in provincial area	0.2445	0.1984***
Attends school in remote area	0.0126	0.0040***
n	258290	213163

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: t-tests are conducted for Ho: proportion of students with given characteristic for private school = proportion of public students with said characteristic, against the two-tailed alternative. Thus, significance is only reported for the private school proportions.

Table 1 illustrates the differences in student background between public and private schools. It can be seen that, for private schools, both the mother and the father are significantly more likely to have completed year 12, attained a bachelor's degree, and to be currently employed. Furthermore, a student in a private school is less likely to come from an Indigenous or language other than English background. Clearly, students in private schools have a significant socio-educational advantage over their peers. These variables are therefore also all included in the model. We also include dummies for the state the student resides in, and a year dummy to control for differences in testing period.

It is also important to note that these sample statistics highlight how significant a threat to identification selection bias is in our estimation. We would like to isolate the effects of different student characteristics in a private school versus a public school, so it is crucial to understand if students are more likely to succeed in private school or if it is simply the case that students in a private school have some kind of pre-existing socio-educational advantage, on average. These features of the data drive the control function methodology of our paper.

We must also acknowledge that in Australia not only 'advantaged' students attend private schools, and 'disadvantaged' students public. This is especially the case in the younger years of schooling where many students attend public schools before transferring to the private system for high school. The Australian context provides good variance in student background features in both schooling systems, making it ideal for comparatively analysing disadvantage in the public and private sectors. As can be seen in Table C2 in Appendix C, there is good variation in regressors for each schooling sector and both sectors have students from both significantly advantaged and disadvantaged backgrounds.

3 Theoretical model

We model educational achievement using the education production function approach, as discussed by Hanushek (1979). Educational achievement, a_i^*, is a function of a set of inputs, $\boldsymbol{x}_i$:

$$a_i^* = f(\boldsymbol{x}_i) \tag{1}$$

These inputs are commonly school resources, other school features, and student background characteristics. In many existing studies, school resources have been shown to, on average, have no systematic relationship with outcomes (Hanushek, 2003),[13] so can be omitted from empirical estimations of production functions without introducing substantial omitted variable bias.

We are left with two determinants of outcomes, school-level factors and student background characteristics. Student background is determined by characteristics such as parental education and employment, gender, and race. School characteristics often include community factors such as location, organisational features such as public or private status, and peer features such as average cohort test scores. We model educational achievement as a function of the following inputs; organisation (public or private), mother and father's education levels and employment status, Indigenous status, language background other than English, gender, and school location (state/territory of Australia, and rural/metropolitan status). We model latent educational achievement, a_i^*, as the outcome a_{ji}, an ordered categorical variable that represents levels of achievement on a standardised test.

Consider the following potential outcomes framework for educational achievement, a_{ji}:

$$a_{ji} = \boldsymbol{x}'_{ji}\beta_j + D'_{ji}\theta_j + v_{ji}, \quad j = 0, 1; i = 1, \ldots, n. \tag{2}$$

$$D_{ji} = \begin{cases} 0, & \text{if } \boldsymbol{z}'_{ji}\gamma_j + u_{ji} \leq 0 \\ 1, & \text{if } \boldsymbol{z}'_{ji}\gamma_j + u_{ji} > 0 \end{cases} \tag{3}$$

[13]Reference the preceding chapter for further discussions on this. While we found evidence of a relationship between test scores and funding and certain points of the distribution, it was not found that there is a relationship on average, that is, when running a simple OLS model.

where $j = 0$ for public schooling and $j = 1$ for private schooling. D_{ji} represents potential outcomes for a student, where D_{0i} gives their potential outcome in public schooling and D_{1i} the outcome of the same student in private schooling. θ_j measures the effect of private schooling on educational achievement, that is, how the probability of meeting a given achievement band varies with type of schooling. The $\boldsymbol{x}_{ji}$ contains weakly exogenous student background covariates, and the $\boldsymbol{z}_{ji}$ include these covariates as well as additional covariates partially correlated with D_{ji} to provide variation separate from that in the achievement equation. β_j therefore measures how the probability of meeting a given achievement band varies with each of the characteristics contained in $\boldsymbol{x}_{ji}$. We have theorised this effect as β_{ji} so that we may understand the effect of these characteristics on educational achievement for students in private schools, $j = 1$, as well as the different effect these characteristics may have students in a public school, $j = 0$, *i.e.* $\beta_{0i} \neq \beta_{1i}$. γ_j measures how the probability of attending private school varies with each of the characteristics contained in $\boldsymbol{z}_{ji}$. The (v_{ji}, u_{ji}) are assumed to follow a bivariate normal distribution with mean 0 and covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. We allow the case that $\rho \neq 0$, to account for self-selection into private schooling, thus the v_{ji} in equation (2) are linearly correlated with the u_{ji} in equation (3).

In reality, each student is observed in only one outcome - public or private schooling. Let $Priv$ denote a dummy variable equal to one if individual i actually receives private schooling and zero if public schooling. Given (2) - (3), the observed outcome, a_i, and the D_i are:

$$a_i = Priv \cdot a_{1i} + (1 - Priv)a_{0i}, \quad i = 1, \dots, n \tag{4}$$

$$D_i = Priv \cdot D_{1i} + (1 - Priv)D_{0i}, \quad i = 1, \dots, n \tag{5}$$

As $\boldsymbol{x}_{ji}$ and $\boldsymbol{z}_{ji}$ contain the same variables for both the public schooling and private school cases, we can substitute (2) and (3) into (4) and (5) respectively, giving the model:

$$a_i^* = \boldsymbol{x}_i'\beta_0 + D_{1i}'\theta_1 + (Priv \cdot \boldsymbol{x}_i')\delta + \varepsilon_i \tag{6}$$

$$D_i = \mathbb{1}[\boldsymbol{z}_i'\gamma + \tilde{u}_i \geq 0] \tag{7}$$

where $\delta = \beta_1 - \beta_0$, $\zeta = \gamma_1 - \gamma_0$, $\varepsilon_i = v_{1i} + (v_{1i} - v_{0i})Priv$, and $\tilde{u}_i = u_{1i} + (u_{1i} - u_{0i})Priv$. We would like to find δ, the difference in the effect of each student background characteristic on educational achievement between public and private school students.

4 Estimation strategy

To account for the correlation between u_{ji} and v_{ji}, we implement the control function approach suggested by Wooldridge (2015). Our method is outlined below.

1. Estimate equation (3) using probit, $\mathbb{P}[D_i = 1|\boldsymbol{z}_i] = \Phi(\boldsymbol{z}_i'\gamma)$, where $\Phi(\cdot)$ is the cumulative density function for a standard normal distribution.

2. Obtain the 'generalised residuals', an estimate for the unobservables contained in u_i. As discussed in Wooldridge (2015), it is well known that under our given assumptions:

$$\mathbb{E}[a_i^*|\boldsymbol{z}_i, D_i] = \boldsymbol{x}_i'\beta + \rho[D_i\lambda(\boldsymbol{z}_i'\gamma) - 1 - (1 - D_i)\lambda(-\boldsymbol{z}_i'\gamma)] \tag{8}$$

where $\lambda(\cdot)$ is the inverse Mills ratio.[14] The function $r(D_i, \boldsymbol{z}_i\gamma) = D_i\lambda(\boldsymbol{z}_i'\gamma) - 1 - (1-D_i)\lambda(-\boldsymbol{z}_i'\gamma)$ is the generalised residual, and has a mean of zero conditional on $\boldsymbol{z}_i$.

3. Estimate equation (2) using maximum likelihood with the addition of these estimated generalised residuals, $\hat{r}_i$:

$$a_i^* = \boldsymbol{x}_i'\beta + D_i'\delta + \hat{r}_i'\rho + \varepsilon_i \tag{9}$$

where the error term, ε_i, is now uncorrelated with the D_i to consistently estimate β, δ, and ρ[15]. a_i^* is not observed, we instead observe $a_i = 1$ if $a_i^* < c_1$, $a_i = 2$ if $c_1 \leq a_i^* < c_2$, or $a_i = 3$ if $a_i^* \geq c_2$. c_1 and c_2 are certain thresholds of achievement, where $c_1 < c_2$. Multinomial ordered probit is therefore used to estimate this second stage.

4. Compute the marginal effects for the multinomial probit at the mean so we can understand the effects for an 'average' student. As most variables are dummy variables, the marginal effect δ for the outcome of not meeting minimum standards, for example, is given by $\mathbb{P}(a_i = 1|D_i = 1, \boldsymbol{x}_i, \hat{r}_i) - \mathbb{P}(a_i = 1|D_i = 0, \boldsymbol{x}_i, \hat{r}_i)$. We then average these effects across all of the data to determine the marginal effect at the mean.

5. Bootstrap the standard errors to correct for the two-stage procedure.

Naive estimation would have assumed that $\rho = 0$ in the covariance matrix of the v_{ji} in equation (2) and u_{ji} in equation (3). This would imply that there is no correlation between equation (2) and equation (3), or, as above, $\rho = 0$. Identification would be based on the assumption that selection into private schooling is based only on observables, in other words, conditional on the covariates in the vector $\boldsymbol{x}_i$, potential outcomes are mean-independent of schooling sector selection. In reality, it is likely that there is unobservable self-selection into attending private school; for example, if these students have some kind of greater underlying ability. The pathway by which this operates is that the propensity to send a

[14] The inverse Mills ratio $\phi(\cdot)/\Phi(\cdot)$ is the same as that typically used in an ordinary Heckman correction.

[15] Consider the t-statistic on the estimate for ρ to simply test the null hypothesis that the private schooling dummy D_i is exogenous, and that the control function correction was appropriate (similarly to the well-known Hausman (1978) test of endogeneity).

child to private school is based on the apparent educational gains that may result from this schooling sector. This selection on unobservables causes the private schooling dummy, D_i to be correlated with the error term, u_i. A simple ordered multinomial probit will obtain upwards biased estimates on the coefficient on the private school dummy, overestimating the effect of private schooling on student achievement.

Given that the model is correctly specified, estimation requires the following assumptions (Wooldridge, 2015; Murtazashvili and Wooldridge, 2016):

ASSUMPTION 1 (Validity). $\mathbb{E}[v_i|z_i] = 0$ *and* $\mathbb{E}[u_i|z_i] = 0$.

ASSUMPTION 2 (Normality). $[u_i, v_i]' \sim N(0, \Sigma)$.

ASSUMPTION 3 (Rank condition). *There is at least one regressor in $\boldsymbol{z}_i$ that is not also in $\boldsymbol{x}_i$ with non-zero coefficient in equation (7).*

Assumption 1 assumes that the instruments contained in $\boldsymbol{z}_i$ are valid. Assumption 2 assumes joint normality of the error terms. Assumption 3 is the usual rank condition requiring at least one instrument partially correlated with a_i.

Following Assumptions 1 and 2, we can write:

$$\mathbb{E}(v_i|\boldsymbol{z}_i, D_i) = \mathbb{E}(v_i|u_j) = \rho u_i, \tag{10}$$

where $\rho = 0$ would correspond to exogeneity of D_i. Substituting, this yields equation (8) where under Assumptions 1 - 3 we can consistently estimate β, δ, and ρ. A simple 2SLS approach would be consistent, but would make no distinction between a binary or continuous D_i. If the assumptions for consistency of the control function method holds, the control function estimator is most efficient (Wooldridge, 2015).

To ensure Assumption 1, we require at least one exogenous variable that causes variation in D_i not explained by $\boldsymbol{x}_i$ to obtain suitable estimates for v_i. We must therefore develop an understanding of what factors are driving school choice in Australia. We assume that parents attempt to choose the 'best' school for their child, but there are numerous factors underlying this decision that may not involve a school's ability to causally improve their child's test scores.[16] For example, a legacy of private schooling in a family may increase the likelihood that those parents will send their child to a private school, and vice versa. A parent may be interested in matching their child to a school based on ethos, religion, or simply from positive recommendations from their neighbours. To further complicate the matter, some families will be restricted in the set of schools that they may choose from depending on the level of fees charged and their child's ability to earn a scholarship. Ultimately, we focus on a condition that appears to drive school choice, but not student outcomes. This condition is the job category of the mother. As shown by Table F4, the proportion of children in the lowest band of achievement who have a mother with a job in categories 3 and 4 remains relatively consistent at just above 20%.[17] It is, therefore, unlikely that having a mother with a job in these categories affects the likelihood of not meeting minimum standards, which is what we are interested in. A job in category 3 refers to tradespeople, clerks, and skilled office, sales, and service staff. Some examples of these jobs are; personal assistants, sales staff, flight attendants, fitness instructors, and child care workers. A job in category 4 covers machine operators, hospitality staff, assistants, labourers, and related workers; for example, office assistants, miners, farmers, guards, and members

[16]Dearden et al. (2011) provides a discussion of the determinants of private school choice in Australia and the United Kingdom.

[17]23.32 and 23.02 respecitively for reading, and 22.68 and 22.89 for numeracy.

of the defence force ranked below senior non-commissioned officer.

While these occupations of the mother do not appear to influence educational achievement, they do indeed affect choice in schooling. In particular, the largest proportion of students attending private school have a mother in job category 3, whereas students in a public school are more likely to have a mother who is not employed. It is intuitive that the mother's employment would determine school choice, as it is the variable contained in our data set which can most tell us something about the financial situation and labour allocation to work of a family, which is likely to be a significant driver of school choice, without impacting achievement. Further analysis of these variable choices can be found in Appendix F.

There are a number of reasons as to why this variable may not satisfy the exclusion restriction, which must be kept in mind. To confirm its validity we would need to analyse the sensitivity of mothers to the labour force to understand how workforce entry/exit may be related to school choice, as well as other choices related to organisation of the family. Census data (at the family unit) is difficult to link to information about the child's school attendance, making it hard to analyse underlying trends about family decision-making for their children. It is indeed possible that the results of this paper may hold best only for particular subgroups of the Australian population. However, given the current form of the data, the use of these variables provides the best attempt at sourcing an instrument, as it does not appear to be correlated with potential confounding channels in the data. This is a research area that using the LSAC or LSAY data could be valuable in strengthening our understanding of family decision-making and school choice. Future versions of this paper will work on using auxiliary data sets to strengthen the analysis.

5 Results

5.1 Private schooling

We analyse the results for each level of the outcome, *educational achievement*, to determine the marginal effect of private schooling for an average student in each achievement band. Recall that as each variable is a dummy variable, the marginal effect can be interpreted as the change in probability of a given outcome occurring by switching that dummy variable from zero to one. For example, $\mathbb{P}[a_{i1}|\boldsymbol{x}_i, D_i = 1] - \mathbb{P}[a_{i1}|\boldsymbol{x}_i, D_i = 0]$ gives the change in probability from attending private school (switching the private schooling dummy from 'off' to 'on') on a given student achieving a NAPLAN score in outcome band 1 (not meeting minimum standards). In the tables, this is displayed as the marginal effect of private schooling. The marginal effects are non-linear for our model, so we therefore evaluate them at the mean of each regressor so that they can be interpreted as the effect for an 'average' student. The estimates for the probit regression, as well as some alternative specifications for the model, can be found in the Appendices. The marginal effects are presented in the following Tables 2 - 4.

What these results show are that, on average, private school students are more likely to achieve above standards on NAPLAN in reading and math, and less likely to achieve below standards than public school students. However, the impact that private schooling has on a student's outcomes depends on their level of achievement. While private schooling has a marginal effect of 0.104 for both reading and numeracy for the outcome 'achieving above standards', the marginal effects are small for those achieving below minimum standards.

The marginal effects for this outcome are only -0.016 and -0.011. These results would indicate that while private schooling may provide gains for students already achieving above minimum standards, school choice matters little for those achieving below standards. Nevertheless, at the margin there may be some gains for students to make. The marginal effects for just meeting minimum standards are -0.088 and -0.093 respectively, meaning that students at this level of achievement are about 8 or 9 percentage points less likely to be in this band of achievement if they attend private school. These results provide some evidence that school matching may enable students at the boundary of poor achievement to meet minimum standards. Potentially, private schooling enables students to lift their scores just enough that they can meet minimum standards. An area for further research is to understand how private schools are either successfully targeting these at risk students, or alternatively if they are 'gaming' the system when sitting NAPLAN tests in some way.[18]

These results also indicate that conditioning on observable demographics alone are not enough to capture selection into the private schooling sector. The coefficients on the predicted residuals are statistically significant at the 1% level in all cases, indicating that, in terms of NAPLAN achievement, there exists self-selection into school sectors based on unobservable characteristics. It is also interesting to note that the marginal effects for achieving above standards increase in magnitude after correcting for endogeneity. If it is indeed true that student ability is driving these school choice decisions, we would expect to see decreased marginal effects following the implementation of the control function approach. What this result indicates is that it is unlikely to be ability that is driv-

[18]See Battistin et al. (2017) for evidence of score manipulation in Italy, and Reback (2008) for an analysis of 'teaching to the test' in the United States.

ing school choice decisions in Australia. For example, it has been suggested that a familial legacy of private schooling is the main determinant of school choice in Australia.[19] It may also be the case that parents with a child of high ability know they will be able to succeed in either a public or private environment, whereas parents of lower ability children may select private schools due to their apparent gains in NAPLAN achievement. The sign on the predicted residuals was in fact negative for the outcome 'achieving above minimum standards', indicating that the unobservables correlated with private schooling and achievement are such that the characteristics they contain disadvantage students in NAPLAN. This would support the story that it is in fact students at the margin of lower ability whose parents choose private schooling.

5.2 Student background and educational achievement

Next, we analyse the role of certain background factors in determining student achievement. The results show that both parental characteristics and student characteristics are statistically significant in determining outcomes. The marginal effects for the background characteristic variables have been computed for each outcome (in the same manner as the preceding section) and presented in Tables 2 - 4 alongside the marginal effects for private schooling.

It can be seen that a student whose parents completed year 12, have Bachelor's degrees and are employed are less likely to achieve below minimum standards on NAPLAN. However, the impact of these variables is generally small. For example, the marginal effect for a mother holding a bachelor's degree qualification is only -0.007. The impact of all of these variables is larger for reading than

[19] See Dearden et al. (2011) for analysis of these other potential determinants of private school choice in Australia.

numeracy; for example, the marginal effect for a mother holding a Bachelor's degree on achieving below standards decreases to -0.004 for numeracy. The effect is over twice as large for reading than for numeracy for many cases. It can therefore be concluded that parent background will affect a student's achievement in mathematics less than in reading.

In terms of student background, there is a comparably large marginal effect for Indigenous status, of 0.009 for reading and 0.011, making these students much more likely to achieve below standards in both reading and mathematics respectively. This indicates that addressing problems involved with Indigenous education remains a significant policy issue in Australia. Secondly, although the effect is small, students with a language background other than English are more likely to score below minimum standards in both reading and mathematics. The effect is smaller for mathematics, which is logical as their language background will have less of an effect in this subject area than in reading, which is tested in English.

It is interesting to note that the marginal effect of being female has opposing signs for reading and numeracy. The marginal effect for achieving below standards is -0.006 for reading, but 0.001 for mathematics. This could indicate that girls have improved ability in reading, or potentially teaching styles for reading in Australia favour female styles of learning, whereas mathematics classrooms may favour a male learning style. Nevertheless, these differences are relatively small in magnitude.

At the school level, location seems to matter to some extent, with students in provincial areas less likely to succeed in both reading and numeracy than a student in the city (the base group). However, these effects are also once again, perhaps surprisingly, small. It appears that after controlling for other student background features, location seems only to have a small effect on the likelihood of a student meeting minimum NAPLAN standards. It is possible that improved results of metropolitan students are therefore a result of the public/private choice of schooling, rather than a pure location effect. This would lead us to conclude that perhaps differences in achievement in Australia are due to differences in opportunities provided to socio-economically advantaged and disadvantaged students, rather than being driven by ability. This would be consistent with our unintuitive increase in the marginal effect once accounting for sample selection, as perhaps it is not unobservable ability driving private school attendance, which is often argued. Further work needs to be done to understand what is driving school choice, as our understanding of this decision is limited here by our data. These results need not mean that all students should be attending private school, but simply indicate that improved outcomes for disadvantaged students are indeed possible. We require an increased policy focus on correcting inequities of opportunity in the Australian school system, to ensure that no student is limited by their socio-economic background.

If we look at the results from the other side of the coin, that is, the marginal effects for a student achieving above standards, we see a similar picture. From Table 4, we can see that a student with a mother and father who completed year 12, hold a Bachelor's degree, and are working are more likely to achieve above standards in both reading and math. However, mother's employment seems to play little role in achievement above minimum standards.

Table 2: Marginal effects of a multinomial probit regression on NAPLAN achievement, corrected for endogeneity (%)

Outcome: not meeting standards	**Reading**	**Numeracy**
Private schooling	−0.0160***	−0.0112***
	(0.0013)	(0.0011)
Parental characteristics		
Mother completed year 12	−0.0059***	−0.0053***
	(0.0002)	(0.0002)
Mother holds a bachelor's degree	−0.0069***	−0.0040***
	(0.0002)	(0.0002)
Mother employed	0.0006***	0.0002
	(0.0002)	(0.0002)
Father completed year 12	−0.0050***	−0.0039***
	(0.0002)	(0.0002)
Father holds a bachelor's degree	−0.0067***	−0.0024***
	(0.0002)	(0.0001)
Father employed	−0.0041***	−0.0046***
	(0.0004)	(0.0003)
Student characteristics		
Indigenous	0.0090***	0.0105***
	(0.0007)	(0.0008)
LBOTE	0.0044***	0.0035***
	(0.0002)	(0.0002)
Female	−0.0057***	0.0011***
	(0.0002)	(0.0001)
Attends school in provincial area	0.0016***	0.0004**
	(0.0002)	(0.0002)
Attends school in remote area	0.0003	0.0002
	(0.0008)	(0.0006)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Marginal effects calculated at the mean. Standard errors in parentheses and constructed using the bootstrap. The sign on $\hat{r}$ is positive and significant.

Table 3: Marginal effects of a multinomial probit regression on NAPLAN achievement, corrected for endogeneity (%)

Outcome: at standard	**Reading**	**Numeracy**
Private schooling	−0.0877***	−0.0931***
	(0.0071)	(0.0093)
Parental characteristics		
Mother completed year 12	−0.0306***	−0.0420***
	(0.0009)	(0.0012)
Mother holds a bachelor's degree	−0.0398***	−0.0205***
	(0.0009)	(0.0013)
Mother employed	−0.0032***	−0.0016
	(0.0011)	(0.0014)
Father completed year 12	−0.0266***	−0.0314***
	(0.0010)	(0.0012)
Father holds a bachelor's degree	−0.0386***	−0.0205***
	(0.0010)	(0.0013)
Father employed	−0.0208***	−0.0344***
	(0.0018)	(0.0022)
Student characteristics		
Indigenous	0.0418***	0.0696***
	(0.0029)	(0.0040)
LBOTE	0.0227***	0.0273***
	(0.0011)	(0.0014)
Female	−0.0311***	0.0094***
	(0.0008)	(0.0010)
Attends school in provincial area	0.0088***	0.0029**
	(0.0010)	(0.0012)
Attends school in remote area	0.0019	0.0019
	(0.0043)	(0.0045)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Marginal effects calculated at the mean. Standard errors in parentheses and constructed using the bootstrap. The sign on $\hat{r}$ is positive and significant.

Table 4: Marginal effects of a multinomial probit regression on NAPLAN achievement, corrected for endogeneity (%)

Outcome: above standards	**Reading**	**Numeracy**
Private schooling	0.1037***	0.1043***
	(0.0085)	(0.0104)
Parental characteristics		
Mother completed year 12	0.0363***	0.0473***
	(0.0011)	(0.0014)
Mother holds a bachelor's degree	0.0467***	0.0382***
	(0.0011)	(0.0014)
Mother employed	−0.0038***	−0.0018
	(0.0013)	(0.0016)
Father completed year 12	0.0316***	0.0352***
	(0.0012)	(0.0014)
Father holds a bachelor's degree	0.0452***	0.0229***
	(0.0012)	(0.0014)
Father employed	0.0249***	0.0390***
	(0.0021)	(0.0025)
Student characteristics		
Indigenous	−0.0507***	−0.0801***
	(0.0036)	(0.0047)
LBOTE	−0.0271***	−0.0308***
	(0.0013)	(0.0016)
Female	0.0368***	−0.0105***
	(0.0009)	(0.0012)
Attends school in provincial area	−0.0104***	−0.0032**
	(0.0012)	(0.0014)
Attends school in remote area	−0.0022	−0.0022
	(0.0051)	(0.0055)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Marginal effects calculated at the mean. Standard errors in parentheses and constructed using the bootstrap. The sign on $\hat{r}$ is negative and significant.

6 Model predictions

In the model above, we have developed a tool for predicting NAPLAN achievement of individual students based on socio-educational background determinants. In this section we analyse whether a student's achievement can accurately be predicted using this model. We generate the probabilities for students with certain background features to reach certain levels of achievement, and see how these probabilities change when certain characteristics are 'switched on' or 'off'. We are interested in how the probability of meeting certain levels of achievement changes with public or private schooling, for students with different background characteristics. Models like these may have significant ability to predict poor performance before it occurs, thus enabling policy-makers, schools, and teachers to provide attention where it is most required.

Specifically, we calculate the predicted probabilities for each type of student by controlling their background characteristics. Recall that the marginal effects and predicted probabilities for the probit model are non-linear, and thus will change by manipulating what values we input to the model. For example, to predict the probability of a disadvantaged female student in a public school achieving above standards, we substitute $Private = 0$, $girl = 1$, $Indigenous = 1$, *etc.*, into the model in the preceding section to calculate that type of student's likelihood of achieving above standards on NAPLAN. Given the results of the previous section, we define 'advantaged' characteristics as any characteristic that has a positive sign for surpassing minimum standards. That is, an advantaged student has a mother and father who completed year 12 and a Bachelor's degree, and is employed. They do not come from an Indigenous or LBOTE background and attend school in the city. A 'disadvantaged' student is the opposite of this. We then vary (*i.e.*, set to either zero or one) certain other characteristics such as language background, living in a provincial area, Indigenous background, and gender.

The following table presents the predicted probabilities for achieving below the minimum standard on NAPLAN testing, given certain sets of characteristics. The predicted probabilities for achieving above minimum standards are reproduced in Table 6 for the same sets of background characteristics.

Table 5: Predicted probabilities for not meeting minimum standards, for certain socio-educational groups

Background characteristics			Reading Public	Reading Private	Numeracy Public	Numeracy Private
Advantaged	Female	-	0.0025***	0.0003***	0.0064***	0.0018***
			(0.0002)	(0.0000)	(0.0005)	(0.0001)
Advantaged	Male	-	0.0047***	0.0007***	0.0056***	0.0016***
			(0.0004)	(0.0001)	(0.0004)	(0.0001)
Disadvantaged	Female	-	0.0899***	0.0263***	0.0862***	0.0373***
			(0.0050)	(0.0041)	(0.0043)	(0.0050)
Disadvantaged	Male	-	0.1295***	0.0421***	0.0797***	0.0340***
			(0.0065)	(0.0062)	(0.0040)	(0.0047)
Advantaged	Female	LBOTE	0.0038***	0.0006***	0.0088***	0.0026***
			(0.0003)	(0.0001)	(0.0006)	(0.0002)
Advantaged	Male	LBOTE	0.0071***	0.0011***	0.0079***	0.0023***
			(0.0005)	(0.0001)	(0.0005)	(0.0002)
Advantaged	Female	Indigenous	0.0052***	0.0008***	0.0136***	0.0043***
			(0.0004)	(0.0001)	(0.0009)	(0.0004)
Advantaged	Male	Indigenous	0.0094***	0.0016***	0.0122***	0.0038***
			(0.0007)	(0.0002)	(0.0008)	(0.0004)
Advantaged	Female	Provincial	0.0029***	0.0004***	0.0066***	0.0019***
			(0.0002)	(0.0000)	(0.0004)	(0.0002)
Advantaged	Male	Provincial	0.0055***	0.0008***	0.0059***	0.0017***
			(0.0004)	(0.0001)	(0.0004)	(0.0001)
Disadvantaged	Female	Provincial	0.0976***	0.0292***	0.0868***	0.0377***
			(0.0033)	(0.0035)	(0.0028)	(0.0041)
Disadvantaged	Male	Provincial	0.1395***	0.0465***	0.0804***	0.0343***
			(0.0040)	(0.0052)	(0.0026)	(0.0038)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses and constructed using the bootstrap.

These predicted probabilities show wide differences in the likelihood that certain students will meet given levels of achievement, based off background factors alone. They also indicate the potential for gains to private schooling for certain types of student, and not for others. For example, advantaged male and female students have high probabilities of achieving above minimum standards

Table 6: Predicted probabilities for achieving above minimum standards, for certain socio-educational groups

Background characteristics			Reading		Numeracy	
			Public	Private	Public	Private
Advantaged	Female	-	0.9625***	0.9913***	0.8720***	0.9399***
			(0.0021)	(0.0006)	(0.0050)	(0.0027)
Advantaged	Male	-	0.9415***	0.9848***	0.8806***	0.9447***
			(0.0029)	(0.0010)	(0.0047)	(0.0025)
Disadvantaged	Female	-	0.6209***	0.8174***	0.3034***	0.6650***
			(0.0116)	(0.0177)	(0.0108)	(0.0224)
Disadvantaged	Male	-	0.5380***	0.7558***	0.5201***	0.6802***
			(0.0122)	(0.0213)	(0.0107)	(0.0220)
Advantaged	Female	LBOTE	0.9487***	0.9871***	0.8452***	0.9242***
			(0.0027)	(0.0009)	(0.0053)	(0.0036)
Advantaged	Male	LBOTE	0.9221***	0.9782***	0.8550***	0.9300***
			(0.0036)	(0.0015)	(0.0050)	(0.0034)
Advantaged	Female	Indigenous	0.9368***	0.9832***	0.8028***	0.8979***
			(0.0035)	(0.0014)	(0.0069)	(0.0054)
Advantaged	Male	Indigenous	0.9059***	0.9722***	0.8143***	0.9052***
			(0.0046)	(0.0022)	(0.0067)	(0.0052)
Advantaged	Female	Provincial	0.9574***	0.9898***	0.8692***	0.9383***
			(0.0022)	(0.0008)	(0.0047)	(0.0030)
Advantaged	Male	Provincial	0.9343***	0.9824***	0.8780***	0.9433***
			(0.0029)	(0.0013)	(0.0045)	(0.0028)
Disadvantaged	Female	Provincial	0.6032***	0.8049***	0.5017***	0.6635***
			(0.0071)	(0.0143)	(0.0067)	(0.0181)
Disadvantaged	Male	Provincial	0.5196***	0.7411***	0.5184***	0.6787***
			(0.0071)	(0.0170)	(0.0068)	(0.0179)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses and constructed using the bootstrap.

on both their NAPLAN reading and numeracy tests in both public and private schools. In contrast, socio-educationally disadvantaged students are much more likely to succeed in a private school. This is especially the case for female disadvantaged students, who only have a 30% probability of achieving above standards in numeracy if attending a public school, which more than doubles to 66% in a private school.

The results for students with an Indigenous background or language background other than English are also particularly interesting. Students with these backgrounds, who are otherwise advantaged, have much higher probabilities of good performance. These results indicate that these characteristics need not be causing poor performance in and of themselves.

The probabilities for both advantaged and disadvantaged students living in provincial areas have also been reproduced in these tables. The results indicate again that living in a provincial area in itself need not determine poor outcomes, as otherwise advantaged students living in provincial areas have high probabilities of meeting minimum standards regardless of their location. What all of these results would indicate is that it is the parent background factors, specifically their education levels, that are most strongly driving student achievement on NAPLAN. This would be consistent with the literature on intergenerational mobility, which finds education to be a transmission mechanism for higher economic outcomes across generations of a family.[20]

7 Conclusion

Test score data has shown large differences in achievement between students in public and private school systems in Australia. We aim to understand the causes of these differences by analysing what background factors predict poor NAPLAN performance in Australian children. Secondly, we seek to determine if private schooling may shield the effects of disadvantage to enable students with these characteristics to meet minimum educational standards.

We find that there are numerous characteristics that predict poor NAPLAN performance in Australian children. Specifically, children with an Indigenous or language other than English background are at risk of not meeting minimum standards. Furthermore, students with a parent who did not complete year 12,

[20]See Bjorklund and Jantti (1997); Corak (2013); Clark (2014); Chetty, Hendren, Kline and Saez (2014); Mendolia and Siminski (2017) for more of the literature on intergenerational mobility. While this is not the focus of this paper, the results would show that levels of education have a role to play in the transmission of economic outcomes between parents and children. While the literature is inconsistent in the magnitude of the effects of intergenerational transmission of outcomes, as further NAPLAN data becomes available (linked to more types of economic outcome data) more can be done to understand the mechanisms at work.

did not attend university or is not employed are also more likely to perform poorly. Students who attend schools in provincial or remote areas are also at risk of poor NAPLAN performance. In terms of private schooling, we find that there are gains for some particular student groups, but not all. An implication of these results is that perhaps school vouchers to improve student to school matching or expanding scholarship programs for disadvantaged students with particular risk factors to attend private schools may improve student achievement.

There are certain important statements that can be made about educational achievement from these results. Firstly, it appears that disadvantage is generally a cumulative problem. While the effect of having any one of these features is small, if a student has a large number of negative characteristics the effect on their test score will be compounded. In a socio-economically disadvantaged student, these small marginal effects could add up to a significant educational disadvantage rapidly. For example, if a student's mother has not completed Year 12 she will not have a bachelor's degree and may also not be employed. As it is likely that markers of disadvantage exist in concert with each other, the impact of socio-economic background on educational achievement remains a significant problem in Australia. The probabilities predicted by our model suggest that simply manipulating one or two background characteristics leads to large changes in a student's likelihood of achieving minimum standards. This result is positive, as it indicates that there is potential for policy to be targeted more directly to students with particular characteristics, such as Indigenous students, to improve their outcomes. It is well-known that in Australia many disadvantaged members of society suffer from entanglement in the 'web of disadvantage', where the effect of one form of disadvantage can reinforce impacts of other disadvantage markers. These members of society face significant constraints to the opportunities they receive, and more can be done to address these markers directly to diminish their impacts. In this case, disadvantage may restrict the set of schools that parents can choose from for their child, meaning that their child may not be matched to the best school possible for them. Policies aimed at improving school matching should therefore be considered by the government; however, more work needs to be done on a data set containing more covariates (potentially the LSAC) to confirm this finding and better understand the mechanisms at work.

Secondly, these results indicate that parent educational attainment is the most significant driver of a student's achievement in NAPLAN, given the variables in this data set. The policy consideration from this finding is that by encouraging higher levels of educational attainment across society in the long-term we may be able to improve educational standards in Australia for future generations. This conclusion should also be considered alongside future studies using the LSAC, LSAY, or census data in order to understand the mechanisms at work in greater detail.

A NAPLAN score equivalence tables

Table A1: 2013 NAPLAN score equivalence table

	Grade 3		Grade 5		Grade 7		Grade 9	
Band	Reading	Numeracy	Reading	Numeracy	Reading	Numeracy	Reading	Numeracy
1	266	264						
2	322*	314*						
3	369	366	368	371				
4	424	417	419*	422*	419	423		
5	469	475	471	475	477*	477*	471	477
6	771	741	523	529	523	526	528*	526*
7			574	571	575	578	580	580
8			758	834	627	631	630	632
9					890	900	686	683
10							891	977

Note: Starred bands mark the minimum standard for that grade and subject in that year. Scores reported are the upper limit for that band.
Source: acara.edu.au.

Table A2: 2014 NAPLAN score equivalence table

	Grade 3		Grade 5		Grade 7		Grade 9	
Band	Reading	Numeracy	Reading	Numeracy	Reading	Numeracy	Reading	Numeracy
1	269	252						
2	319*	317*						
3	365	369	369	371				
4	420	417	424*	424*	421	424		
5	467	475	475	471	473*	476*	478	473
6	772	750	528	529	526	529	529*	527*
7			577	572	579	580	576	578
8			812	815	631	632	627	631
9					935	886	684	682
10							908	921

Note: Starred bands mark the minimum standard for that grade and subject in that year. Scores reported are the upper limit for that band.
Source: acara.edu.au.

B Variables contained in the NAPLAN student-level data set

Table B1: Variables contained in the student-level NAPLAN data set

Variable	Description
Reading standard	=0 if student is below standard in reading, =1 if at standard, and =2 if above standard
Math standard	=0 if student is below standard in numeracy, =1 if at standard, and =2 if above standard
Private schooling	=1 if student attends a non-government school, and =0 if government school
Age	Age at the time of taking the test to one decimal place
LBOTE	=1 if the student has a language background other than English, and =0 if not
Indigenous	=1 if the student identifies as Indigenous Australian or Torres Strait Islander, and =0 if not
Female	=1 if the student is female, and =0 if male
Metropolitan	=1 if the student lives in a metropolitan location, and =0 if not
Provincial	=1 if the student lives in a provincial location, and =0 if not
Rural	=1 if the student lives in a remote or very remote location, and =0 if not
State dummies	Dummy variables =1 if the student lives in that state (South Australia, New South Wales, Victoria, Queensland, Tasmania, Western Australia, Australian Capital Territory, Northern Territory), and =0 if not
Mother year 12	=1 if the student's mother finished year 12, and =0 if not
Mother Bachelor's degree	=1 if a student's mother's highest level of education is a Bachelor's degree, and =0 if not
Mother diploma	=1 if a student's mother's highest level of education is a Diploma or Advanced Diploma, and =0 if not
Mother certificate	=1 if a student's mother's highest level of education is a certificate I to IV or trade qualification, and =0 if not
Mother not employed	=1 if a student's mother is not employed, and =0 if not
Mother job category 1	=1 if a student's mother employed in category 1,[21] and =0 if not
Mother job category 2	=1 if a student's mother employed in category 2,[22] and =0 if not
Mother job category 3	=1 if a student's mother employed in category 3,[23] and =0 if not
Mother job category 4	=1 if a student's mother employed in category 4,[24] and =0 if not
Father year 12	=1 if the student's mother finished year 12, and =0 if not
Father Bachelor's degree	=1 if a student's father's highest level of education is a Bachelor's degree, and =0 if not
Father diploma	=1 if a student's father's highest level of education is a Diploma or Advanced Diploma, and =0 if not
Father certificate	=1 if a student's father's highest level of education is a certificate I to IV or trade qualification, and =0 if not
Father not employed	=1 if a student's father is not employed, and =0 if not
Father job category 1	=1 if a student's father employed in category 1, and =0 if not
Father job category 2	=1 if a student's father employed in category 2, and =0 if not
Father job category 3	=1 if a student's father employed in category 3, and =0 if not
Father job category 4	=1 if a student's father employed in category 4, and =0 if not

Note: Raw data provided by the Australian Curriculum, Assessment and Reporting Authority through their Data Access Program. Our expanded dataset also includes the interactions and squared terms of each variable (excluding the band of achievement variables and the private schooling variable as these are outcome variables in the main equation and selection equation). Interactions are also not included for the state dummies.

[21] Senior management in a large business organisation, government administration, or defence, and qualified professionals; *e.g.,* business/policy analyst, defence forces commissioned officer, professionals with degree or higher qualifications, administrators such as school principals, *etc.*

[22] Other business managers, arts/media/sportspersons, and associate professionals; *e.g.,* owner/manager of a farm or business, retail sales/service manager, musician, journalist, designer, sports official, business/administrative staff, *etc.*

[23] Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*

[24] Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*

C Supplementary data

Table C1: Proportion of students at each standard of achievement by schooling sector

	Public schools			**Private schools**		
	Above	At	Below	Above	At	Below
Reading	0.8702	0.1113	0.0184	0.9068***	0.0827***	0.0105***
Numeracy	0.8136	0.1716	0.0148	0.8431***	0.1482***	0.0120***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: n=471453. 'Above' refers to the proportion of students above standard, 'At' refers to students just meeting the minimum standard, and 'Below' refers to students not meeting minimum standards. t-tests are conducted for Ho: proportion of students in a given band for private school = proportion of public students in said band, against the two-tailed alternative. Thus, significance is only reported for the private school proportions.

Table C2: Standard deviation of background characteristic regressors by schooling sector

	Public	**Private**
Parental characteristics		
Mother completed year 12	0.482	0.444
Mother holds a bachelor's degree	0.495	0.496
Mother employed	0.453	0.380
Father completed year 12	0.498	0.472
Father holds a bachelor's degree	0.481	0.500
Father employed	0.277	0.190
Student characteristics		
Indigenous	0.163	0.102
LBOTE	0.426	0.406
Female	0.500	0.500
Attends school in city	0.437	0.402
Attends school in provincial area	0.430	0.399
Attends school in remote area	0.112	0.063

D Alternative probit specifications

Table D1: Ordered probit estimates of the determinants of student achievement

	(1)		(2)		(3)		(4)	
	Reading	**Numeracy**	**Reading**	**Numeracy**	**Reading**	**Numeracy**	**Reading**	**Numeracy**
Private school	0.184***	0.109***	0.068***	0.022**	0.160***	0.092***	0.056***	0.014
	(0.015)	(0.012)	(0.012)	(0.012)	(0.015)	(0.012)	(0.012)	(0.011)
Parental characteristics								
Mother completed year 12			0.213***	0.192***			0.215***	0.194***
			(0.006)	(0.006)			(0.006)	(0.006)
Mother holds a bachelor's degree			0.305***	0.172***			0.304***	0.169***
			(0.008)	(0.008)			(0.008)	(0.008)
Mother employed			0.045***	0.049***			0.027***	0.032***
			(0.007)	(0.006)			(0.007)	(0.032)
Father completed year 12			0.200***	0.154***			0.207***	0.161***
			(0.006)	(0.006)			(0.006)	(0.006)
Father holds a bachelor's degree			0.313***	0.117***			0.312***	0.117***
			(0.008)	(0.010)			(0.008)	(0.009)
Father employed			0.223***	0.222***			0.177***	0.180***
			(0.013)	(0.012)			(0.012)	(0.012)
Student characteristics								
Indigenous					−0.474***	−0.460***	−0.305***	−0.324***
					(0.017)	(0.016)	(0.012)	(0.016)
LBOTE					−0.122***	−0.109***	−0.168***	−0.135***
					(0.014)	(0.016)	(0.011)	(0.012)
Female					0.208***	−0.037***	0.218***	−0.038***
					(0.007)	(0.006)	(0.006)	(0.006)
Attends school in provincial area					−0.176***	−0.100***	−0.077***	−0.027***
					(0.015)	(0.014)	(0.014)	(0.013)
Attends school in remote area					−0.241***	−0.174***	−0.143***	−0.108***
					(0.033)	(0.029)	(0.033)	(0.028)
State dummies	yes	yes	yes	yes	yes	yes	yes	yes
Observations	471453	471453	471453	471453	471453	471453	471453	471453
Pseudo R2	0.0074	0.0085	0.0586	0.0331	0.0188	0.0128	0.0671	0.0358
log pseudo-L	-900994	-1130960	-854463	-1102893	-890631	-1126038	-846835	-1099826
$Pr >$ Wald χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Standard errors in parentheses and clustered by school.

E Marginal effects for a naive ordered multinomial probit regression on NAPLAN achievement

Table E1: Marginal effects of a multinomial probit regression on NAPLAN achievement

Outcome: not meeting standards	**Reading**	**Numeracy**
Private schooling	−0.002***	−0.001***
	(0.000)	(0.000)
Parental characteristics		
Mother completed year 12	−0.006***	−0.006***
	(0.000)	(0.000)
Mother holds a bachelor's degree	−0.007***	−0.004***
	(0.000)	(0.000)
Mother employed	−0.001***	−0.001***
	(0.000)	(0.000)
Father completed year 12	−0.006***	−0.004***
	(0.000)	(0.000)
Father holds a bachelor's degree	−0.007***	−0.003***
	(0.000)	(0.000)
Father employed	−0.006***	−0.006***
	(0.000)	(0.000)
Student characteristics		
Indigenous	0.012***	0.012***
	(0.001)	(0.001)
LBOTE	0.005***	0.004***
	(0.000)	(0.000)
Female	−0.006***	0.001***
	(0.000)	(0.000)
Attends school in provincial area	0.002***	0.001***
	(0.000)	(0.000)
Attends school in remote area	0.005***	0.003***
	(0.001)	(0.001)

Table E2: Marginal effects of a multinomial probit regression on NAPLAN achievement

Outcome: at standard	**Reading**	**Numeracy**
Private schooling	−0.009***	−0.004*
	(0.001)	(0.003)
Parental characteristics		
Mother completed year 12	−0.033***	−0.044***
	(0.001)	(0.001)
Mother holds a bachelor's degree	−0.042***	−0.037***
	(0.001)	(0.002)
Mother employed	−0.003***	−0.007***
	(0.001)	(0.001)
Father completed year 12	−0.031***	−0.036***
	(0.001)	(0.001)
Father holds a bachelor's degree	−0.042***	−0.025***
	(0.001)	(0.002)
Father employed	−0.028***	−0.043***
	(0.002)	(0.003)
Student characteristics		
Indigenous	0.052***	0.080***
	(0.003)	(0.004)
LBOTE	0.026***	0.031***
	(0.002)	(0.003)
Female	−0.032***	0.008***
	(0.001)	(0.001)
Attends school in provincial area	0.012***	0.006*
	(0.002)	(0.003)
Attends school in remote area	0.023***	0.025***
	(0.006)	(0.007)

Table E3: Marginal effects of a multinomial probit regression on NAPLAN achievement

Outcome: above standards	**Reading**	**Numeracy**
Private schooling	0.010***	0.005***
	(0.001)	(0.001)
Parental characteristics		
Mother completed year 12	0.0392***	0.050***
	(0.001)	(0.002)
Mother holds a bachelor's degree	0.049***	0.041***
	(0.001)	(0.002)
Mother employed	0.005***	0.008***
	(0.001)	(0.002)
Father completed year 12	0.0368***	0.041***
	(0.001)	(0.001)
Father holds a bachelor's degree	0.050***	0.028***
	(0.001)	(0.002)
Father employed	0.034***	0.0484***
	(0.003)	(0.003)
Student characteristics		
Indigenous	−0.063***	−0.093***
	(0.004)	(0.005)
LBOTE	−0.031***	−0.035***
	(0.002)	(0.003)
Female	0.0377***	−0.009***
	(0.001)	(0.002)
Attends school in provincial area	−0.014***	−0.007**
	(0.002)	(0.003)
Attends school in remote area	−0.027***	−0.028***
	(0.007)	(0.008)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: Marginal effects calculated at the mean. Standard errors in parentheses and clustered by school.

F Choice of first-stage variables

Table F1: Basic probit models for first and second stages

Dependent variable	**Private schooling**	**Reading band**	**Numeracy band**
Potential instruments			
Mother with a category 3 job	−0.143***	−0.043***	−0.021***
	(0.005)	(0.007)	(0.006)
Mother with a category 4 job	−0.448***	−0.154***	−0.132***
	(0.006)	(0.008)	(0.007)
Parental characteristics			
Mother completed year 12	0.060***	0.207***	0.188***
	(0.005)	(0.006)	(0.005)
Mother holds a bachelor's degree	−0.030***	0.263***	0.129***
	(0.005)	(0.008)	(0.006)
Mother employed	0.471***	0.091***	0.069***
	(0.005)	(0.007)	(0.006)
Father completed year 12	0.153***	0.204***	0.159***
	(0.005)	(0.006)	(0.005)
Father holds a bachelor's degree	0.130***	0.304***	0.097***
	(0.005)	(0.008)	(0.006)
Father employed	0.345***	0.187***	0.188***
	(0.008)	(0.009)	(0.008)
Student characteristics			
Indigenous	−0.379***	−0.305***	−0.320***
	(0.015)	(0.014)	(0.014)
LBOTE	−0.088***	−0.156***	−0.129***
	(0.005)	(0.006)	(0.005)
Female	0.031***	0.221***	−0.035***
	(0.004)	(0.005)	(0.004)
Attends school in provincial area	−0.120***	−0.074***	−0.022***
	(0.005)	(0.006)	(0.005)
Attends school in remote area	−0.636***	−0.111***	−0.043***
	(0.023)	(0.024)	(0.022)
State dummies	yes	yes	yes
Year dummies	yes	yes	yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: For the private schooling outcome a probit model was estimated. Multinomial ordered probits were estimated for the reading band and numeracy band outcomes. Coefficients are therefore not interpretable as marginal effects. Standard errors in parentheses. A job in category three refers to "Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*". A job in category four refers to "Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*".

Table F2: 2SLS model instrumenting for private schooling

	Reading	**Numeracy**
Private schooling	0.214***	0.229***
	(0.011)	(0.012)
Parental characteristics		
Mother completed year 12	0.052***	0.055***
	(0.001)	(0.002)
Mother holds a bachelor's degree	0.039***	0.030***
	(0.001)	(0.002)
Mother employed	−0.016***	−0.017***
	(0.002)	(0.002)
Father completed year 12	0.039***	0.034***
	(0.002)	(0.002)
Father holds a bachelor's degree	0.030***	0.009***
	(0.002)	(0.002)
Father employed	0.036***	0.042***
	(0.003)	(0.003)
Student characteristics		
Indigenous	−0.076***	−0.092***
	(0.004)	(0.005)
LBOTE	−0.026***	−0.028***
	(0.001)	(0.002)
Female	0.044***	−0.013***
	(0.001)	(0.001)
Attends school in provincial area	−0.007***	0.004***
	(0.001)	(0.002)
Attends school in remote area	0.016**	0.035***
	(0.006)	(0.007)
State dummies	yes	yes
Year dummies	yes	yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Standard errors in parentheses. Instruments used are; *Mother with a job in category 3*, and *Mother with a job in category 4*. A job in category three refers to "Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*". A job in category four refers to "Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*".

Table F3: First stage of 2SLS model

Private schooling	
Instruments	
Mother with a category 3 job	−0.057***
	(0.002)
Mother with a category 4 job	−0.173***
	(0.002)
Parental characteristics	
Mother completed year 12	0.021***
	(0.002)
Mother holds a bachelor's degree	0.012***
	(0.002)
Mother employed	0.180***
	(0.002)
Father completed year 12	0.058***
	(0.002)
Father holds a bachelor's degree	0.051***
	(0.002)
Father employed	0.118***
	(0.003)
Student characteristics	
Indigenous	−0.123***
	(0.005)
LBOTE	−0.034***
	(0.002)
Female	0.012***
	(0.001)
Attends school in provincial area	−0.045***
	(0.002)
Attends school in remote area	−0.213***
	(0.008)
State dummies	yes
Year dummies	yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Standard errors in parentheses. A job in category three refers to "Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*". A job in category four refers to "Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*".

Table F4: Tabulation for the proportions of students with a mother in each occupational group across schooling sectors and achievement bands (%)

	Category 1	Category 2	Category 3	Category 4	Not employed	Total
Reading band						
Below standards	7.33	14.16	23.32	23.02	32.18	100.00
At standards	8.98	16.15	25.06	20.56	29.26	100.00
Above standards	20.52	20.88	22.44	13.16	23.00	100.00
Numeracy band						
Below standards	6.81	13.93	22.68	22.89	33.70	100.00
At standards	14.35	17.13	23.41	17.99	27.11	100.00
Above standards	20.31	21.03	22.58	13.14	22.95	100.00
Schooling sector						
Public schooling	14.45	17.59	21.94	17.08	28.94	100.00
Private schooling	24.93	23.62	23.64	10.35	17.37	100.00

Note: Explanations of each job category are given in Appendix B.

G Ordered probit model corrected for endogeneity

Table G1: Estimates for an ordered probit model corrected for endogeneity

	Reading	Numeracy
Private schooling	0.598***	0.418***
	(0.048)	(0.042)
$\hat{r}$	−0.331***	−0.248***
	(0.029)	(0.025)
Parental characteristics		
Mother completed year 12	0.201***	0.184***
	(0.006)	(0.005)
Mother holds a bachelor's degree	0.288***	0.158***
	(0.007)	(0.006)
Mother employed	−0.022***	−0.007
	(0.007)	(0.006)
Father completed year 12	0.178***	0.140***
	(0.007)	(0.005)
Father holds a bachelor's degree	0.281***	0.094***
	(0.008)	(0.006)
Father employed	0.133***	0.147***
	(0.011)	(0.009)
Student characteristics		
Indigenous	−0.252***	−0.284***
	(0.016)	(0.015)
LBOTE	−0.148***	−0.120***
	(0.007)	(0.006)
Female	0.212***	−0.042***
	(0.005)	(0.005)
Attends school in provincial area	−0.059***	−0.013**
	(0.007)	(0.006)
Attends school in remote area	−0.013	−0.009
	(0.029)	(0.022)
State dummies	yes	yes
Year dummies	yes	yes

* p < 0.1, ** p < 0.05, *** p < 0.01.

Note: Standard errors in parentheses and corrected using the bootstrap. Instruments used are; *Mother with a job in category 3*, and *Mother with a job in category 4*. A job in category three refers to "Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*". A job in category four refers to "Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*".

Table G2: Estimates of a probit model for private schooling (Control function first-stage)

Private schooling	
Instruments	
Mother with a category 3 job	−0.142***
	(0.005)
Mother with a category 4 job	−0.438***
	(0.006)
Parental characteristics	
Mother completed year 12	0.048***
	(0.005)
Mother holds a bachelor's degree	−0.026
	(0.005)
Mother employed	0.457***
	(0.005)
Father completed year 12	0.143***
	(0.005)
Father holds a bachelor's degree	0.136***
	(0.005)
Father employed	0.290***
	(0.008)
Student characteristics	
Indigenous	−0.353***
	(0.014)
LBOTE	−0.071***
	(0.005)
Female	0.031***
	(0.004)
Attends school in provincial area	−0.088***
	(0.005)
Attends school in remote area	−0.829***
	(0.023)
State dummies	yes
Year dummies	yes

* p < 0.1, ** p < 0.05, *** p < 0.01.

Note: Heteroskedasticity-robust errors in parentheses. A job in category three refers to "Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*". A job in category four refers to "Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*".

H Sensitivity to data subset

Table H1: Marginal effects of a multinomial probit regression on NAPLAN achievement (10% stratified sample, $n = 47144$)

Outcome: not meeting standards	**Reading**	**Numeracy**
Private schooling	−0.010**	−0.006*
	(0.005)	(0.004)
Parental characteristics		
Mother completed year 12	−0.006***	−0.005***
	(0.001)	(0.001)
Mother holds a bachelor's degree	−0.007***	−0.005***
	(0.001)	(0.000)
Mother employed	0.000	0.000
	(0.001)	(0.001)
Father completed year 12	−0.006***	−0.004***
	(0.001)	(0.001)
Father holds a bachelor's degree	−0.008***	−0.003***
	(0.001)	(0.000)
Father employed	−0.003***	−0.004***
	(0.001)	(0.001)
Student characteristics		
Indigenous	0.012***	0.011***
	(0.003)	(0.002)
LBOTE	0.006***	0.004***
	(0.001)	(0.001)
Female	−0.006***	0.002***
	(0.001)	(0.000)
Attends school in provincial area	0.003***	0.001**
	(0.001)	(0.001)
Attends school in remote area	0.004	0.003
	(0.003)	(0.003)

Table H2: Marginal effects of a multinomial probit regression on NAPLAN achievement (10% stratified sample, $n = 47144$)

Outcome: at standard	**Reading**	**Numeracy**
Private schooling	−0.050**	−0.050*
	(0.024)	(0.031)
Parental characteristics		
Mother completed year 12	−0.030***	−0.041***
	(0.003)	(0.004)
Mother holds a bachelor's degree	−0.038***	−0.042***
	(0.003)	(0.004)
Mother employed	−0.001	−0.002
	(0.003)	(0.005)
Father completed year 12	−0.028***	−0.032***
	(0.003)	(0.004)
Father holds a bachelor's degree	−0.044***	−0.027***
	(0.004)	(0.004)
Father employed	−0.017***	−0.030***
	(0.006)	(0.007)
Student characteristics		
Indigenous	0.052***	0.070***
	(0.009)	(0.012)
LBOTE	0.029***	0.028***
	(0.003)	(0.004)
Female	−0.032***	0.013***
	(0.002)	(0.003)
Attends school in provincial area	0.013***	0.009**
	(0.003)	(0.004)
Attends school in remote area	0.017	0.022
	(0.014)	(0.019)

Table H3: Marginal effects of a multinomial probit regression on NAPLAN achievement (10% stratified sample, $n = 47144$)

Outcome: above standards	**Reading**	**Numeracy**
Private schooling	0.060**	0.056*
	(0.029)	(0.034)
Parental characteristics		
Mother completed year 12	0.036***	0.046***
	(0.004)	(0.004)
Mother holds a bachelor's degree	0.046***	0.047***
	(0.004)	(0.005)
Mother employed	0.001	0.002
	(0.004)	(0.005)
Father completed year 12	0.034***	0.036***
	(0.003)	(0.005)
Father holds a bachelor's degree	0.051***	0.031***
	(0.004)	(0.005)
Father employed	0.020***	0.034***
	(0.007)	(0.008)
Student characteristics		
Indigenous	−0.064***	−0.081***
	(0.012)	(0.014)
LBOTE	−0.035***	−0.032***
	(0.004)	(0.005)
Female	0.038***	−0.014***
	(0.003)	(0.004)
Attends school in provincial area	−0.015***	−0.010**
	(0.004)	(0.005)
Attends school in remote area	−0.021	−0.025
	(0.016)	(0.022)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Marginal effects calculated at the mean. Standard errors in parentheses and calculated using the bootstrap.

Statement of Authorship

Does machine learning improve predictions of academic achievement? Evidence from Australia
A modified version of this paper has been submitted for publication.

Principal Author

Name: Sarah Cornell-Farrow
Contribution: Idea development, data acquisition and cleaning, program execution (authorised user of data set on this project), drafting and revision of the manuscript.
Overall percentage (%): 70%
Certification: This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

25 January 2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate to include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name: Robert Garrard
Contribution: Technical programming work, manuscript drafting (particularly of Section 4), and editing.

25 January 2019

Does machine learning improve predictions of academic achievement? Evidence from Australia*

Sarah Cornell-Farrow

The University of Adelaide, School of Economics, Adelaide, SA, Australia, 5005.

sarah.cornell-farrow@adelaide.edu.au

Robert Garrard

CSIRO, Land & Water, Dutton Park, QLD, Australia, 4102.

robert.garrard@csiro.au

Abstract

Machine learning methods tend to outperform traditional statistical models at prediction, largely because of their ability to model non-linearities tractably and exploit the bias-variance trade-off. In the prediction of academic achievement, machine learning models have not shown substantial improvement over linear regression and logistic classification. So far, these results have almost entirely focused on college achievement, due to the availability of administrative datasets, and have contained relatively small sample sizes by ML standards. In this paper we apply popular machine learning models to a large dataset ($n = 2.2$ million) containing primary and middle school performance on NAPLAN. We show that machine learning models do not outperform logistic regression for detecting students who will perform in the 'below standard' band of achievement upon sitting their next test.

Keywords: Education economics, NAPLAN, machine learning, standardised testing, Australia.

JEL Classification: I20, I29, C45, C55.

*This research is supported by an Australian Government Research Training Program (RTP) Scholarship. Thanks must also be given to the Australian Curriculum, Assessment and Reporting Authority for the provision of the data utilised by this study. The authors would like to thank Foivos Diakogiannis, Airong Zhang, and a number of anonymous referees for their helpful comments and suggestions.

1 Introduction

Not only are a large number of Australian students achieving poor educational outcomes, but there is a significant disparity in achievement levels within the same age groups and even within the same classroom. 'One-size-fits-all' policy approaches have not been successful in bringing under-achieving students up to minimum standards. In order to support these students in reaching their full potential, it may be necessary to tailor policies and teaching strategies to individual students who are at risk of low achievement. Since schools face resource constraints, it is essential to detect 'at risk' students early and with high precision.

In this setting, the goal is not to conduct causal inference on a coefficient of interest, but to produce a purely predictive model that classifies students with high accuracy. Machine learning (ML) methods tend to greatly outperform traditional statistical models. For example, when attempting to classify images of handwritten digits correctly, standard logistic regression achieves an accuracy of around 70%, whereas state-of-the-art neural networks obtain 99.79% accuracy (Wan et al., 2013). Machine learning estimators perform well by exploiting the bias-variance trade-off, in which the estimator is permitted to be biased in exchange for a large reduction in its variance, as well as incorporating non-linearities in a tractable way. It has been argued that machine learning will bring significant change to the way economists analyse policy (Einav and Levin, 2014; Varian, 2014; Mullainathan and Spiess, 2017),[1] and the publication of the NAPLAN data in Australia provides an opportunity to test the performance of

[1] As a practical example, see Ben-David and Frank (2009); Chang et al. (2016); Luo et al. (2017); Bacham and Zhao (2017); Guegan and Hassani (2018); Adams (2018) for more on the finance conversation about utilising machine learning approaches to generate better credit scores.

machine learning in a real-world setting.[2]

The application of machine learning methods to education data has been referred to as Educational Data Mining (Romero and Ventura, 2007). Its use in the prediction of academic performance has been predominantly in the higher education context (Vandamme et al., 2007; Kotsiantis, 2012; Yadav and Pal, 2012; Jishan et al., 2015) largely due to the availability of administrative data sets collected by universities.[3] Interestingly, when machine learning models are benchmarked against logistic regression, they show no or unsubstantial improvement (Kotsiantis et al., 2004; Cortez and Silva, 2008; Huang and Fang, 2013; Gray et al., 2013).

In this paper we exploit a large data set containing scores on the Australian National Assessment Program - Literacy and Numeracy (NAPLAN). This data set contains raw scores for all students who sat the test in the years 2013 and 2014, as well as administrative data on students' individual- and family-level characteristics. In total, the data set contains observations on 2.2 million unique students. As we are interested in detecting below standard achievement, we label students into two classes: 'At Standard' and 'Below Standard', according to whether or not their score meets minimum achievement standards as determined by the Australian Curriculum, Assessment, and Reporting Agency (ACARA). This is done for two learning areas: literacy and numeracy. We split students into those in grade 3, for whom this would be their first time sitting NAPLAN, and students in grades 5 and above, for whom their previous achievement on NAPLAN may be used as a predictor.

We train a set of popular machine learning classifiers with standard logistic

[2]See Pugh and Foster (2014) for a discussion of 'Big Data' in the Australian education setting.

[3]See Shingari et al. (2017) for a recent review.

regression serving as a benchmark. We find that none of the machine learning models outperform logistic regression. It would therefore appear that machine learning methods do not improve our ability to predict poor educational performance. However, machine learning methods, such as the decision tree, may still provide valuable insight to the policy-maker as they can provide simple diagnostic rules for the teacher and the policy-maker. For example, for grades 5 and above the trees show that students who did not previously meet minimum standards in reading are at risk of also falling behind in numeracy in their next test, and vice versa. It is interesting that these trees do not exploit socio-educational status variables, as we would have expected. To the best of our knowledge, this is the first paper to apply machine learning methods to the prediction of primary and middle school student achievement in a large sample size setting.

The rest of this paper proceeds as follows. Section 2 describes the data set. Section 3 will give a summary of ML methods, compared to typical regression methods. Section 4 describes pre-processing of the data and the particular methods used in this paper to build the set of classifiers. Section 5 presents the results, and Section 6 concludes.

2 Data

The National Assessment Program - Literacy and Numeracy (NAPLAN)[4] is a set of standardised literacy and numeracy tests sat by all students in Australia in Years 3, 5, 7, and 9, in both the government and non-government schooling sectors. NAP is described as providing "the measure through which governments, education authorities and schools can determine whether or not young Australi-

[4]See www.nap.edu.au/naplan

ans are meeting important educational outcomes."[5] The tests cover five learning areas known as 'test domains': Reading, Writing, Spelling, Grammar and Punctuation, and Numeracy. The tests are designed to assess student performance relative to the Australian curriculum, at a minimum standards level.

We use individual student-level data from the NAPLAN reading and numeracy tests administered in 2013 and 2014. The NAPLAN reading test involves a reading comprehension-style test on a range of texts, including imaginative, persuasive, and informative. The questions are designed to test knowledge and interpretation of English language in context. The NAPLAN numeracy tests assess students on their performance in mathematics, namely number and algebra, measurement and geometry, and statistics and probability.

The data set contains 2,235,804 unique student IDs, who are attending 9,250 different schools in both the public and private sectors in all states across Australia. NAPLAN scores are calculated to be comparable across testing cycles, so we can combine the 2013 and 2014 data sets into a pooled panel data set and consider these test scores as the students' score in the 'current' time period. All individuals in the data set are unique, as a student sitting NAPLAN testing in 2013 would not sit the test in 2014, and vice versa. The data set therefore covers all Australian students who were in grades 2-9 in the calendar year 2013, although their actual years of sitting the test may differ. The individual scores for each student in each domain are collected by the Australian Curriculum Assessment and Reporting Authority (ACARA), alongside student background information which is collected by schools from students' parents or carers via enrolment forms. Table 1 provides a detailed summary of the variables contained in the data set.

[5] See nap.edu.au

Table 1: Variables contained in the student-level NAPLAN data set

Variable	Description
Reading standard	=1 if student is below standard in reading, and =0 if above standard
Math standard	=1 if student is below standard in numeracy, and =0 if above standard
Private schooling	=1 if student attends a non-government school [*priv1*], and =0 if government school
Age	Age at the time of taking the test to one decimal place [*age*]
LBOTE	=1 if the student has a language background other than English [*lbote1*], and =0 if not
Indigenous	=1 if the student identifies as Indigenous Australian or Torres Strait Islander [*indig1*], and =0 if not
Female	=1 if the student is female [*girl1*], and =0 if male
Metropolitan	=1 if the student lives in a metropolitan location [*met1*], and =0 if not
Provincial	=1 if the student lives in a provincial location [*provincial1*], and =0 if not
Remote	=1 if the student lives in a remote or very remote location [*remote1*], and =0 if not
State dummies	Dummy variables =1 if the student lives in that state (South Australia [*SA1*], New South Wales [*NSW1*], Victoria [*VIC1*], Queensland [*QLD1*], Tasmania [*TAS1*], Western Australia [*WA1*], Australian Capital Territory [*ACT1*], Northern Territory [*NT1*]), and =0 if not
Mother's education	Mother completed up to Grade 9 [*mumschool1*], Grade 10 [*mumschool2*], Grade 11 [*mumschool3*], or Grade 12 [*mumschool4*]
Mother's higher education	Mother's highest level of education is a certificate I to IV or trade qualification [*mumhighed5*], a Diploma or Advanced Diploma [*mumhighed6*], Bachelor's degree [*mumhighed7*], or none [*mumhighed8*]
Mother's employment	Mother employed in category 1 [*mumoccup1*]†, category 2 [*mumoccup2*]‡, category 3 [*mumoccup3*]††, category 4 [*mumoccup4*]‡‡, or is not employed [*mumoccup8*]
Father's education	Father completed up to Grade 9 [*dadschool1*], Grade 10 [*dadschool2*], Grade 11 [*dadschool3*], or Grade 12 [*dadschool4*]
Father's higher education	Father's highest level of education is a certificate I to IV or trade qualification [*dadhighed5*], a Diploma or Advanced Diploma [*dadhighed6*], Bachelor's degree [*dadhighed7*], or none [*dadhighed8*]
Father's employment	Father employed in category 1 [*dadoccup1*]†, category 2 [*dadoccup2*]‡, category 3 [*dadoccup3*]††, category 4 [*dadoccup4*]‡‡, or is not employed [*dadoccup8*]

Note: Raw data provided by the Australian Curriculum, Assessment and Reporting Authority through their Data Access Program. Variable name codes given in square brackets.
†Category 1: Senior management in a large business organisation, government administration, or defence, and qualified professionals; *e.g.,* business/policy analyst, defence forces commissioned officer, professionals with degree or higher qualifications, administrators such as school principals, *etc.*
‡Category 2: Other business managers, arts/media/sportspersons, and associate professionals; *e.g.,* owner/manager of a farm or business, retail sales/service manager, musician, journalist, designer, sports official, business/administrative staff, *etc.*
††Category 3: Tradespeople, clerks, and skilled office, sales, and service staff; *e.g.,* 4 year trade certificate by apprenticeship, clerks, personal assistants, sales, flight attendants, fitness instructors, child care workers, *etc.*
‡‡Category 4: Machine operators, hospitality staff, assistants, labourers and related workers; *e.g.,* machine operators, drivers, labourers, office assistants, defence forces ranked below senior non-commissioned officer, miners, farmers, factory hands, guards, *etc.*

We use these student background observables to detect which students may be at risk of low academic performance so that policies may be directly targeted at the individual level to support their learning. For each testing domain in each year, ACARA determines achievement bands to classify a student's level of achievement based on what particular skills they can perform, *e.g.,*, addition

Table 2: Tabulations for NAPLAN achievement

	Reading	Numeracy
Below minimum standards	364,733	389,676
(%)	(16.29)	(17.40)
Above minimum standards	1,874,321	1,849,378
(%)	(83.71)	(82.60)

of simple numbers, understanding of probability, *etc.*. For each year level, students in the lowest two bands of achievement are deemed to be achieving 'below minimum standards'. Given the raw scores for each student available in the data, we have classified each student into their relevant achievement band following the cut-off scores published by ACARA on their website.[6] Using these bands, we have created a categorical variable which equals zero if a student is meeting minimum standards, and one if they are not, for reading and numeracy respectively.

The tabulations for these variables are presented in Table 2. In the period studied, 364,733 students are not meeting minimum standards in reading and 389,676 in numeracy. This is equivalent to 16.29% and 17.40% of the sample not meeting minimum standards.

3 Summary of machine learning approaches

When undertaking regression, the researcher may have one of two goals: causal inference or prediction. Typically, the economist has focused on the goal of causal inference; however, with the growth of big data, prediction problems have become increasingly of interest. Kleinberg et al. (2015) have argued that these predictive problems are also valuable to the economist, depending on the type of question we wish to answer. In the case of this paper, we are interested

[6] See nap.edu.au

in successfully predicting poor educational performance. This section will summarise machine learning methods in the context of those utilised by this paper, using Friedman et al. (2001) as a guide.[7]

Generally, the objective of regression is to estimate some function that will tell you the expected value of this outcome variable, y, given the values of the observed variables in the data, x. For example, where ε gives some random error:

$$y = m(x) + \varepsilon \tag{1}$$

$$m = \mathbb{E}[y|x], \tag{2}$$

where (2) gives the conditional expectation function, m. Typically, the applied economist will make the simplifying assumption that m is linear in order to estimate this function, giving the usual:

$$y = X\beta + \varepsilon. \tag{3}$$

When estimating, for example, Ordinary Least Squares (OLS), the researcher minimises the sum of squared residuals of a set of data, $||y - \hat{y}||_2^2$. This sum of squared residuals, $\frac{1}{n}||y - \hat{y}||_2^2$,[8] is made up of three parts:

1. irreducible error, $\frac{1}{n}||\varepsilon||_2^2$, which converges to the variance of the random disturbance,

2. bias squared, $\frac{1}{n}||m(x) - \mathbb{E}[\hat{y}]||_2^2$, which gives how far the expected value of the predicted y is from the true value of y, on average; and,

[7] This text provides a useful entry point to machine learning methods for the interested reader.

[8] $||\mathbf{a}||_2^2 = \sum a_i^2$ denotes the ℓ_2 vector norm.

3. variance, $\frac{1}{n}||\hat{y}-\mathbb{E}[\hat{y}]||_2^2$, giving how often the predicted y is above or below the expected value of the predicted y.

Under the Gauss-Markov theorem, OLS gives the best linear unbiased estimator (BLUE) of β if the true CEF is linear and the model is 'correct', *i.e.*, $m(x) = X\beta$ and $\mathbb{E}[X'\varepsilon] = 0$. 'Best' means the variance is lower than that of any other linear unbiased estimator, and unbiased means that the bias is equal to zero, or $\mathbb{E}[\hat{y}] = m(x)$. In other words, the mean squared error of the estimator is as small as possible, with bias set to zero and variance minimised. The unbiasedness of the estimator is typically useful for inference;[9] however, may not be as important for prediction. In order to predict well, we want to minimise this mean squared error on new data $(x_{\text{new}}, y_{\text{new}})$. It is possible that a biased estimator may have a lower mean squared error than the BLUE, trading bias for a smaller variance. This idea is illustrated in Figure 1. The point marked with a circle on the 'bias' line could be illustrative of an OLS estimator, where the bias is low (in fact, zero if the Gauss-Markov assumptions hold). However, if we follow the grey dotted line it can be seen that the variance is relatively high, and the overall predictive error is not minimised for this point. It can also be noted that the OLS estimator lies to the higher end of the model complexity axis. This is because the OLS estimator is relatively multidimensional, as all coefficients take a non-zero value with probability 1. It is indeed possible that there exists an optimal predictor, shown on the diagram with a diamond, such as a penalised least squares estimator. Following the dotted grey line for this point, it can be seen that the bias is higher than that of the OLS estimator; however, the variance

[9] Specifically, the sampling distribution of the estimator should be centred at the true value asymptotically in order to construct confidence intervals with correct coverage. The Lasso, for example, is a consistent estimator whose sampling distribution is asymptotically biased, resulting in difficulty in constructing confidence intervals (Knight and Fu, 2000). However, the Lasso often predicts better than OLS by exploiting the bias-variance tradeoff.

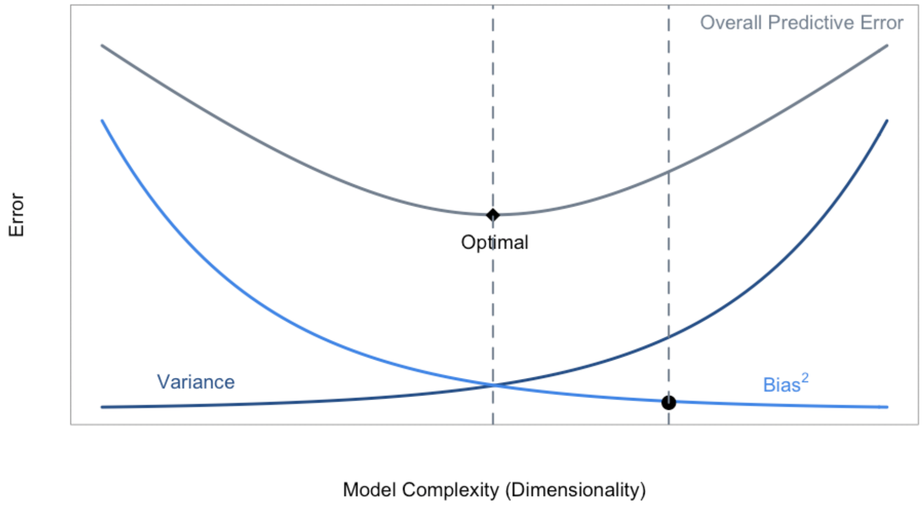


Figure 1: Illustration of the bias-variance trade-off

is lower. Overall, this optimal predictor minimises the predictive error. Thus this biased estimator may predict 'better' on new data than the OLS estimator.

Machine learning uses tuning parameters, sometimes called hyperparameters, to exploit this trade-off between bias and variance. It also allows $m(x)$ to be non-linear. Specifically, we will be focusing on *supervised* learning, where we attempt to predict the value of a dependent variable or outcome Y given a number of independent variables or predictors, X.[10] Furthermore, we will focus on regression for classification for the purposes of this study, that is, where the outcome variable is discrete. Our task is therefore to make a good prediction, $\hat{Y} \in \{0,1\}$, of $Y \in \{0,1\}$. In context, Y may be either *'student is at risk'* for poor educational performance, where $Y = 1$, or *'student is not at risk'*, where $Y = 0$.

A basic machine learning approach to modelling is described in Table 3. To make a prediction, a 'training' set of data is fed into a learning algorithm in order to observe the phenomenon being studied, and fit a function to it. The

[10]This differs from *unsupervised* learning, which aims to describe relationships and patterns between variables, without an outcome Y.

program produces outputs, in our case classifying students as at risk or not at risk. It then 'learns' by comparing the simulated outputs to the actual outputs in the data, modifying the functional relationship between inputs and outputs until the artificial outputs are close enough to reality. This fitting of the model could include both variable selection and parameter estimation. Once the program has 'learnt' about the relationship between inputs and outputs, new data can be fed in to predict future outcomes.

Table 3: Overview of a basic machine learning approach

Algorithm:
1: Split data into a training set, a validation set, and a test set.*
2: Fit initial models to the training data set using a supervised learning method (generate candidate algorithms).
3: Use the fitted model to successively predict the responses for the observations in the validation data set. The validation data tunes the hyperparameters and evaluates the fit of the candidate models.
5: Use the test data set to evaluate overall model performance, *e.g.* sensitivity and specificity. This step also confirms that the model has not been overfitted to the test set, if it performs well on new data.

Note: *A different approach is the use of cross-validation, where the data is split into random test and training subsets a number of times. It is important that the validation set is independent of the data the model was trained on. In this paper we use cross-validation to tune the hyperparameters, as it is allows us to keep the final training set as large as possible.

3.1 Baseline classification: logistic

A baseline classification method for a binary outcome could be as simple as using a logit or probit with no tuning parameters. For example, in the preceding chapter we used the probit model, albeit with causal inference in mind. However, these methods can also be used for prediction when employed on a training set and a test set. Using the training set of data, parameters are picked to maximise the log-likelihood function using maximum likelihood:

$$\ln \mathscr{L} = \sum_{i=1}^{N} \{y_i \ln F(x_i'\beta) + (1 - y_i) \ln F(-x_i'\beta)\}. \tag{4}$$

Most often for prediction, $F(\cdot)$ is the logistic function. Under certain conditions, the maximum likelihood estimator is consistent. This method, however, is a rudimentary effort at allowing non-linearities in the modelling approach, employing a simple monotone transformation to the linear model. As explained above in the case of OLS, it is also true for the logit that there may exist a biased estimator with a smaller mean squared error, which would be better for prediction. In order to improve the predictive accuracy, we could add a tuning parameter to exploit the bias-variance trade-off.

3.2 Ridge regression, the lasso, and elastic net

Next, we look at using penalties for model shrinkage and variable selection. The goal is to make the model more parsimonious than the logit, shifting us toward the 'optimal' predictor. For these models, we trade off some positive bias to decrease the variance and overall prediction error.

First we discuss **ridge regression** (Hoerl and Kennard, 1970), which imposes a penalty on the size of the coefficients to shrink them towards zero and each other. The ridge coefficients minimise a penalised residual sum of squares:

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \tag{5}$$

where $\lambda \geq 0$ controls the amount of shrinkage. Zero is the case of no shrinkage, and the amount of shrinkage increases as λ increases. In a normal linear regression model, if there are many highly correlated variables the coefficients may be poorly determined. A very high positive coefficient on one variable may be cancelled out by a large negative coefficient on another highly correlated variable. The benefit of ridge regression is that by shrinking the coefficients towards zero, this problem can be avoided. Note that, due to the scaling, it is necessary to standardise the variables before implementing ridge regression and to leave the intercept out of the penalty term.

The **lasso estimate** (Tibshirani, 1996) is a similar shrinkage method, minimising the following:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \tag{6}$$

The lasso differs from ridge regression in that an ℓ_1 lasso penalty is used instead of the ℓ_2 ridge penalty term. Instead of shrinking all the coefficients proportionally, the lasso translates each coefficient by the constant factor λ, setting some to zero. A limitation of the lasso method is that if two variables are highly correlated, it tends to arbitrarily set one to zero, giving the full coefficient value to the other. In contrast, ridge would give both of these variables equal weighting.

In this study, we use an **elastic net penalty** (Zou and Hastie, 2005), which aims to find a compromise between these two approaches. This method exhibits the variable selection properties of the lasso, while shrinking together the remaining coefficients of correlated predictors as in ridge regression. This methodology will be discussed in further detail in the methods section of this paper, Section 4.

3.3 Decision trees and random forests

While these approaches are useful for prediction, they may be difficult to interpret in a practical sense. We therefore also implement a decision tree algorithm, as well as a random forest. This algorithm is a variant of the decision tree approach that currently represents the best 'off-the-shelf' machine learning technique. A number of studies have found that decision trees predict well in the education context (Kovacic, 2010; Jishan et al., 2015; Elakia and Aarthi, 2014; Natek and Zwilling, 2014; Mishra et al., 2014).

A **decision tree**, or classification tree, is an algorithm that searches through the data to find conditional statements that split the data in two, along many nodes. A tree can be described as a visual representation as a number of 'decision rules'. These rules that the form: IF *condition1* AND *condition2* THEN *outcome1*. For example, a basic decision tree could look as follows:

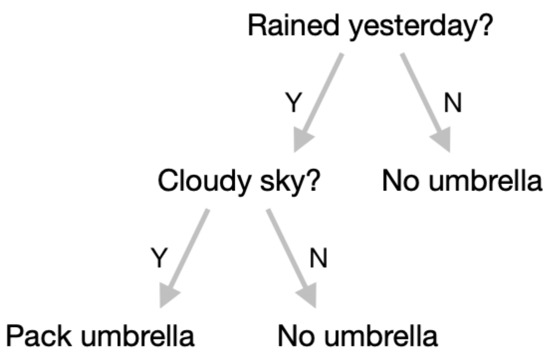


Figure 2: Illustration of a decision tree. Idea based on the prediction problem described in Kleinberg et al. (2015)

.

By following each split of the tree (based on the given predictors), one can determine the outcome of interest. The algorithm for generating a tree on the training set (*i.e.*, step 2 of the overall machine learning algorithm above) may be as follows:

Table 4: Generating a decision tree with 2 predictors, x_1 and x_2

Algorithm:
1: Searches through x_1 and x_2 for the point with the best split (knowing that it will take an average to minimise mean squared error).
2: Takes the average in each region to generate the predicted value.
3: Makes a rule, *e.g.* if $x_1 < 20$.
5: Iterates 1-3 to make another split until some optimal stopping point. Stopping rules may include, for example, preventing a split when the number of observed individuals in a branch reaches some size smaller than a specified percentage of the overall training data.

Note: *N.B* that decision tree processes may also include 'pruning' of its branches to avoid overfitting and improve predictive accuracy. This process removes parts of the tree that provide little power for classification.

Decision trees have the benefit of being easily interpretable, as they provide easy rules to follow for classification in the real world. However, the algorithm is 'greedy', in that it makes a local optimal choice at each stage. This may mean that the final predictor is not the overall optimal predictor. It is possible to improve the accuracy of this process by bootstrap aggregating (or 'bagging') the trees;[11] however, each tree generated will be very similar and thus may not reduce variance. An alternative approach to reduce both the bias and variance of a decision tree approach is to generate a random forest of decision trees.

Random forests (Breiman, 2001) work by growing many single decision trees. For each tree, we sample N observations with replacement from the original data to construct the training set. A random subset, m,[12] of the variables is also selected for each tree. The forest of trees is then grown, without any prun-

[11] A special case of model averaging. See Breiman (1996) for more.

[12] Friedman et al. (2001) describes a rule of thumb for selecting m: for classification purposes, the default value is $\sqrt{p}$ and the minimum node size is one. The best values for m will change dependent on the problem, and thus should be thought of as tuning parameters in practice.

ing. The models are then averaged across, as with bagging. However, unlike the bagging approach, these trees will have a lower correlation with each other due to the omission of some variables from each. The strength of each individual tree, and the correlation between trees, will be reduced by reducing the number of variables, m, selected. The performance of the classifier is therefore sensitive to the choice of m. This method is extremely accurate among existing off-the-shelf algorithms; however, it does not provide the interpretability of the simple decision rules that the single classification tree method gives.

4 Methodology

Our objective is to predict, based on observables in the data set, whether or not a student will perform in the 'below standard' band for their age group upon sitting their next NAPLAN. Note that since family-level student characteristics were collected by the schools themselves, it is reasonable to assume that individual schools will have access to the predictors used in this study well before its students are scheduled to sit their next NAPLAN.

Each row in the data set contains a predictor corresponding to that student's score in reading and numeracy for the NAPLAN most recently sat by that student; *i.e.*, two years previously. Since NAPLAN is compulsory for all students in a given year level, every observation of a student in grades 5 and above has this data. However, since grade 3 is the first grade in which NAPLAN is sat, grade 3 students have missing data for this column.

We split the data set into students in grade 3, and students in grades 5 and above. Classification for the grade 3 students will need to rely on family-level observables; whereas classification for students in grades 5 and up may exploit the presumably strong predictor of previous NAPLAN scores.

4.1 Pre-processing

To obtain the subset of observations containing grades 5 and above, we remove from the sample any rows of observations containing missing data. Every grade 3 student has missing data for the previous reading and math scores, since NAPLAN is first sat in grade 3, and so they are removed from the sample. This reduces the sample size from 2,239,054 observations to 886,392 observations.

In order to construct the subset containing only grade 3 students, we first extract the rows corresponding to students in grade 3. We then delete the columns corresponding to the student's grade, which contains no variation, and the student's previous NAPLAN scores, which are missing. Finally, we remove any rows containing missing data. The final sample for grade 3 students contains 345,817 observations.

For each data set, and for each response variable corresponding to the reading and numeracy standards respectively, we use stratified sampling of the classes to obtain a two thirds/one third split for training and test sets. [13]

4.2 Class Imbalance

The data used in this study exhibits slightly imbalanced class labels, with 1 in 5 observations classified as 'below standard'. Machine learning approaches to classification typically involve fitting a model that minimises a convex loss function that treats observations symmetrically, in that true (false) positives and true (false) negatives are all given equal weight. However, if class imbalance is

[13] Note that since cross-validation will be used to select tuning parameters, we do not employ a validation set.

present in the data such that one class is relatively rare, then symmetric treatment of these observations may lead the loss function to be minimised by classifying all observations as the majority class. Adopting alternate loss functions that treat observations asymmetrically, such as an F-measure (Chinchor and Sundheim, 1993), can lead to non-convexity of the loss function, which raises difficulties in training the classifier. Two common approaches for addressing class imbalance while preserving convexity of the loss function are *weighting* and *re-sampling*.

4.2.1 Weighting

Weighting the observations can be achieved with a loss function of the form

$$\mathscr{L}(\hat{y}, y) = \sum_i L(\hat{y}_i, y_i) w(\hat{y}_i, y_i) \tag{7}$$

where $L(\cdot)$ is a loss function that treats observations symmetrically, y_i is the true class of observation i, $\hat{y}_i$ is the predicted class, and $w(\cdot)$ is a function that assigns a weight to each observation i. Classifiers with such loss functions are referred to as 'cost-sensitive classifiers' (Elkan, 2001), as true (false) positives and negatives may incur different weights in the loss function.

For simplicity, we weight each observation inversely proportional to its class frequency. That is,

$$w(y_i, \hat{y}_i) = \frac{1}{2|\{j \,|\, y_j = y_i\}|} \tag{8}$$

where y_i denotes the true class of observation i and $|\cdot|$ denotes set cardinality. These weights sum to unity.

4.2.2 Re-sampling

An alternative approach is to use an unweighted loss function and approximately balance class instances by resampling from the data. The data may either be under-sampled by randomly choosing a subset of the majority class, or over-sampled by creating random copies of the minority class (Japkowicz et al., 2000). While each of these procedures produce a more balanced data set, under-sampling discards potentially useful information about the majority classes, while over-sampling can tend to overfit noise in the minority classes. 'Synthetic Minority Over-sampling Technique' (Chawla et al., 2002) provides a method of over-sampling that may reduce this overfitting. Rather than over-sampling by directly copying existing observations, SMOTE creates new 'synthetic' observations by choosing points in the feature space which are convex combinations of minority class observations. A combination of over-sampling the minority class using synthetic data together with under-sampling the majority class may then be used. SMOTE sampling may also be used for constructing ensemble classifiers, such as through bootstrap aggregating (bagging) and boosting (Chawla et al., 2003).

We use the package **DMwR** (Torgo and Torgo, 2013) to construct a SMOTE sample which combines synthetic over-sampling of the minority class and under-sampling of the majority class. Ideally, optimal rates for over and under-sampling would be tuned, such as has been considered by Agrawal and Menzies (2018). However, to the best of our knowledge, no implementation for SMOTE tuning is available in R at the time of writing (July, 2018). As such, we use the out-of-the-box parameters for generating the SMOTE sample. For each minority observation, its 5-nearest neighbours are used to generate two additional minority samples; and the majority class is under-sampled to achieve approximate class balance with the synthetic sample.

We generate the SMOTE training sets from the original training sets described in the previous section. We therefore have two training sets - a weighted set and a SMOTE set - for each subject and grade level of interest, *e.g.*, grade 3 Reading. After running each of the models for that subject and grade level, they are all evaluated for performance on the same test set as generated in the data pre-processing.

It is not known *a priori* which correction to the class imbalance problem will perform best, so we train a set of classifiers using both methods. Since the class imbalance is not severe, we focus on weighted classifiers for the results of this paper. The SMOTE results may be found in the Appendices.

4.3 Classifiers

For each data set and class type we estimate four classifiers: once using the full data set combined with observation weights, and then again using the synthetic SMOTE data set.

We first use a logistic classifier and an elastic net (Zou and Hastie, 2005). The loss function takes the form:

$$\mathscr{L}(y,\hat{y}) = l(y,\hat{y}|w,\beta) + \lambda\left(\alpha||\beta||_1 + (1-\alpha)||\beta||_2^2\right) \tag{9}$$

where $l(\cdot)$ denotes the binary cross-entropy loss function[14] with vector of weights w and estimated coefficient vector β; and λ is a tuning parameter determining the strength of the combination ℓ_1 (lasso) and ℓ_2 (ridge) penalty. We impose that the $\alpha = 0.5$, such that the lasso and ridge penalties have equal weight and select

[14] *N.B.* the negative of a logistic log-likelihood function.

λ through 10-fold cross-validation using the R package **glmnet** (Friedman et al., 2009).[15]

A classification tree is trained using the package **rpart** (Therneau et al., 2015), which grows the tree using a recursive partitioning algorithm and prunes it using 10-fold cross-validation.

Finally, we estimate a random forest with an ensemble of 200 trees using the **ranger** package (Wright et al., 2018).

4.4 Evaluation of methods

The output of a classifier is usually a set of predicted 'probabilities', or scores, for each class representing how confident the model is that an observation belongs to that class. The class with the highest score is usually selected to be the predicted class of the observation. In the two class setting, this corresponds to selecting whichever class has predicted probability greater than half. Alternatively, one could choose some desired threshold and classify an observation into the positive class if its predicted probability exceeds that threshold.

To measure classification performance we can analyse two characteristics: sensitivity and specificity. Sensitivity in our case refers to the percentage of students that are at risk of poor performance that are correctly identified, or the true positive rate. Conversely, specificity measures the percentage of students that are not at risk of poor performance that are correctly identified as not being at risk, or the true negative rate. We can choose the desired sensitivity of a classifier by setting the threshold at which observations are classified in the positive class appropriately; although this trades off specificity. The set of sens-

[15]The mixing parameter, α, may also be tuned. However, this option is not included in the glmnet package. We re-estimated the classifiers with $\alpha = 0.1$ and $\alpha = 0.9$ with no significant change to the results.

itivity/specificity pairs that may be achieved by varying this threshold between 0 and 1 can be illustrated by a Receiver Operating Characteristic (ROC) curve. A classifier with a larger area under the ROC curve (referred to as AUC) than another is able to achieve a greater sensitivity for a given specificity, and vice versa. We use AUC to rank the performance of the classifiers.

5 Results

The AUC for each classifier using weighted observations is displayed in Table 5.[16]

Table 5: Performance metrics (AUC) for weighted observations

Grade 3				
Classifier	Literacy		Numeracy	
	AUC	95% CI	AUC	95% CI
Logistic	0.722	(0.717, 0.726)	0.707	(0.702, 0.712)
Elastic Net	0.721	(0.717, 0.726)	0.707	(0.702, 0.712)
Decision Tree	0.656	(0.650, 0.660)	0.668	(0.663, 0.673)
Random Forest	0.687	(0.692, 0.686)	0.681	(0.676, 0.686)
Grade 5+				
Classifier	Literacy		Numeracy	
	AUC	95% CI	AUC	95% CI
Logistic	0.839	(0.837, 0.841)	0.833	(0.830, 0.835)
Elastic Net	0.839	(0.837, 0.841)	0.832	(0.830, 0.835)
Decision Tree	0.767	(0.765, 0.770)	0.760	(0.758, 0.763)
Random Forest	0.829	(0.826, 0.831)	0.823	(0.820, 0.825)

Note: Confidence intervals for AUC are constructed according to Bamber (1975).

In terms of AUC, there is no significant difference in performance between the logistic and the elastic net for all subjects and grades. It is likely that the penalty used to exploit the trade-off between bias and variance is having little effect as the sample size is so large. Moreover, the decision trees and random

[16]The metrics for the SMOTE sample are found in Appendix A. The ROC curves can be found in Appendix B.

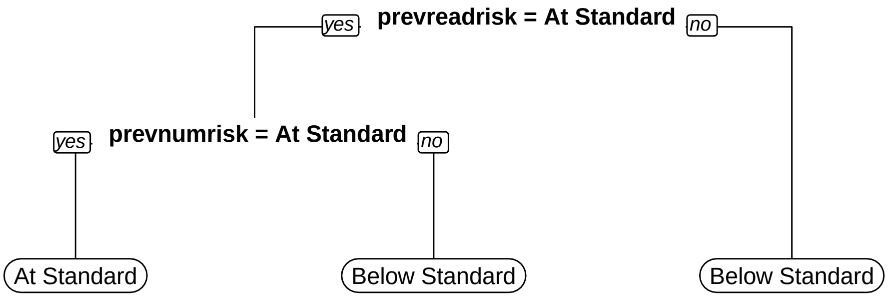
forests perform poorly compared to the simple logistic classifier, indicating that machine learning may not be as useful as some may believe in improving predictions of this type. Non-linearities may not play a large role on the basis of these regressors, as was the case in the higher-education setting.

Nevertheless, the decision trees may be useful in terms of interpretation, as they provide simple rules or heuristics for detecting students at risk of poor performance. For example, the decision trees for grades 5 and above using the weighted observations are shown in Figure 3.[17]

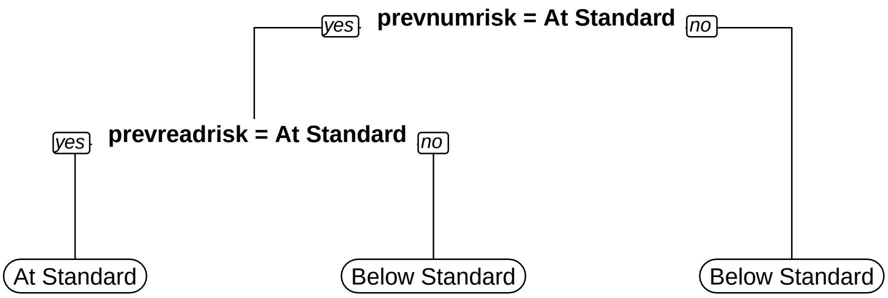
When previous achievement on NAPLAN is available, the decision trees learn a very simple and intuitive decision rule. If a student was below standard on *either* literacy or numeracy in the previous testing period, they are classified as at risk in *both* subjects in the next period. This is of value, as it indicates that a school faced with a student that has performed below standard in numeracy in one period should not shift all of their resources towards improving only their math ability. In the United States, Reback (2008) highlighted that Texan schools re-allocated resources towards students based on their performance on previous tests. We provide evidence that this may not be a useful approach, and instead a focus on overall improvement in all subject areas should be emphasised. This is a particularly interesting finding in the context of "teaching to the test", a popular argument against standardised testing. Perhaps if a student fell behind in reading, for example, in the previous period, so many resources were allocated toward training the student to improve in the reading test that they then fall behind in numeracy in the following period. This is an important area for further research.

The decision trees for grade 3, for which achievement in the previous period

[17] Decision trees for the SMOTE data are reproduced in Appendix C.



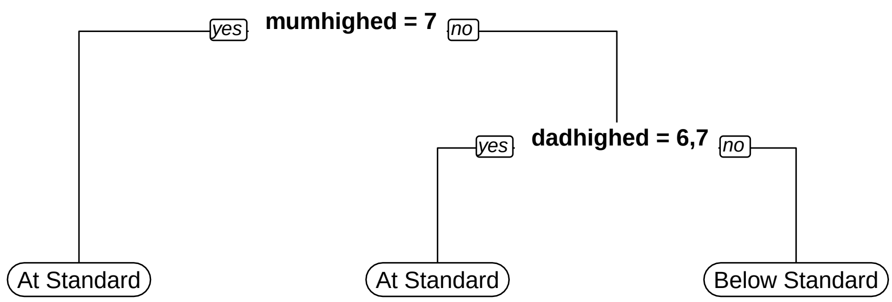
(a) Literacy



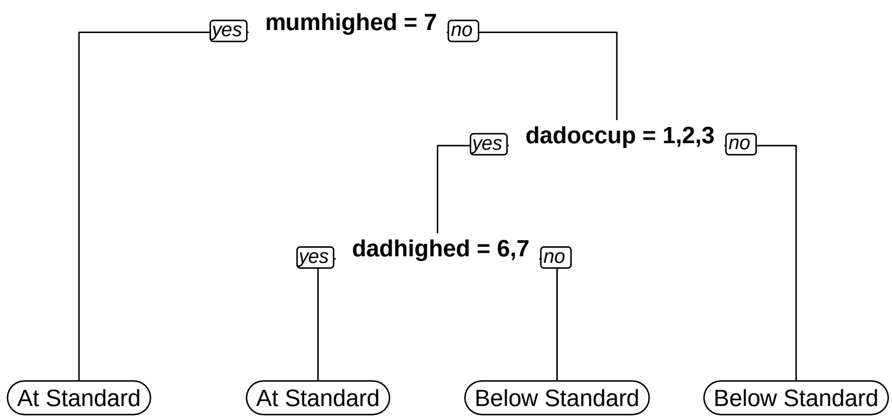
(b) Numeracy

Figure 3: Decison trees for Grades 5+ with weighted observations

is not available, also provide interesting insight to the predictors of poor achievement in Australia. The trees using the weighted observations are shown in Figure 4. Categorising students as at risk or not depends completely on the education levels and occupation of their parents. For both literacy and numeracy, the first split is made regarding mother's higher education. The predictor *mumhighed=7* refers to whether the student's mother holds a Bachelor's degree or not. If a student's mother has a Bachelor's degree, a student will be predicted to



(a) Literacy



(b) Numeracy

Figure 4: Decison trees for Grade 3 with weighted observations.

meet minimum standards. However, if the mother does not hold a degree, their performance depends on the education of their father. If their father also does not have higher education qualifications, the student will be predicted to not meet minimum standards. If their father has either a Bachelor's degree, a diploma or an advanced diploma,[18] the student will be predicted to meet minimum standards. In general terms, if *at least one* of the student's parents has a Bachelor's

[18]If *dadhighed=6*, the student's father has a diploma or advanced diploma.

degree, or their father has a diploma or advanced diploma, a student is not at risk of poor performance on NAPLAN. This finding matches much of the literature, which finds that family background strongly determines educational outcomes (Cobb-Clark and Nguyen, 2012; Ford, 2013). In fact, Nicoletti and Rabe (2013) found that family background explains 44-55% of the variation in test scores for students in England. There could be a number of mechanisms by which this could operate. One example is that perhaps parents with higher levels of education are more likely to actively participate in their child's learning. Fryer et al. (2015) show that incentivising parents to participate in their child's education can improve a student's test scores. As the mother is often the primary caregiver, perhaps a more highly educated mother, who is actively participating in their child's education, enables a child to be more likely to succeed.

The results are the same for literacy and numeracy, save the addition of a decision rule regarding father's occupation for numeracy. *dadoccup=1,2,3* means that the father has a job in either Category 1, 2, or 3. The easiest way to interpret this is that the father is employed, and he is *not* employed in Category 4. Category 4 encompasses machine operators, hospitality staff, assistants, labourers and related workers. Therefore, if a student's father is not employed or employed in one of these types of job, such as a farmer or factory hand, they will be predicted as at risk of not meeting minimum standards in the numeracy test. Behrman and Knowles (1999) found a relationship between a household's income and educational success in Vietnam, so perhaps this decision rule is based on lower household incomes earned in some of these careers. There could be a number of alternate explanations; for example, perhaps if a student has a father in a less math-based career. This is another area for further research.

6 Conclusion

This paper trained a number of machine learning classifiers on the NAPLAN data, in order to find if these techniques can improve the econometrician's ability to predict poor educational performance. It has been posited that machine learning methods will change the way that the applied economist approaches data analysis. The large individual-level NAPLAN data set provides a unique opportunity with which to explore this claim in the context of education.

It has been argued that machine learning methods may be better at predicting outcomes than traditional econometric approaches for two reasons. First, they are able to exploit the trade-off between bias and variance in order to find an estimator with lower predictive error than the unbiased estimator. Second, they can better exploit non-linearities in the data, unlike most typical approaches. In order to test these claims, we estimated an elastic net, a decision tree, and a random forest, and find that these methods fail to out-perform the logistic classifier. The elastic net matched the performance of the logistic, indicating that the penalty used to exploit the bias-variance trade-off did not improve predictive ability. The decision tree and random forests significantly underperformed the logistic and the elastic net. It is therefore unlikely that non-linearities are present, matching findings in the higher education context. These results call into question if machine learning methods can really improve predictive studies as much as some believe.

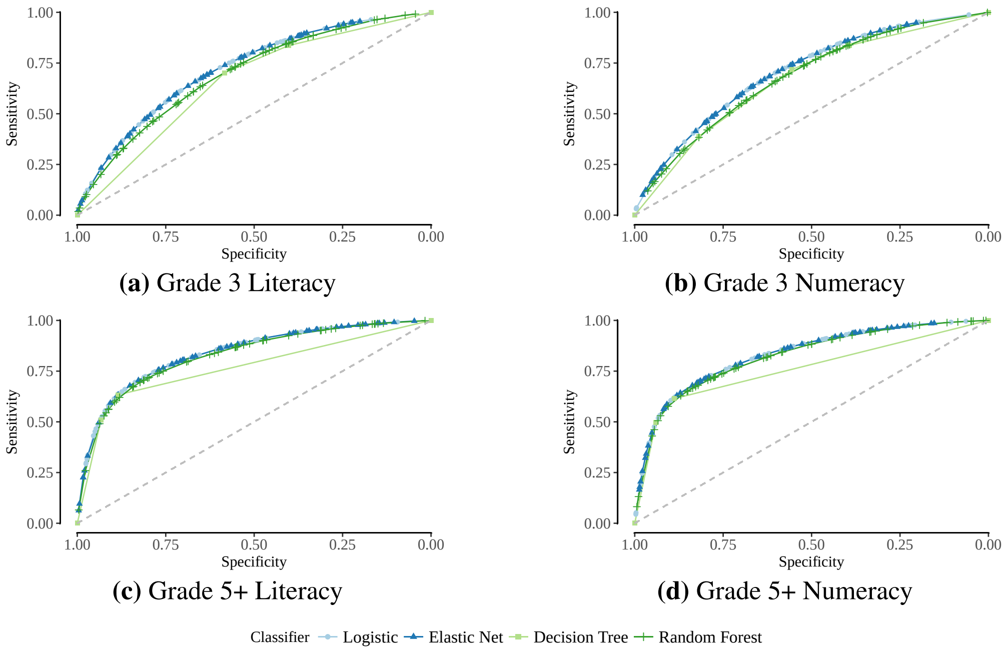
Nevertheless, machine learning methods do provide some insights into the analysis of school education data. A greater emphasis on analysing problems of a predictive type using predictive methods may still provide beneficial. These kinds of analysis can still provide policy insight, despite the focus of the applied economist towards establish causality. Furthermore, they can provide useful heuristics for use in the real world. For example, the decision trees generated in this paper provide simple easy-to-remember rules for application by the school, teacher, and policy-maker. Further discussion is required as to the role of prediction problems in applied economics, but with the caveat that their usefulness not be over-exaggerated.

A Performance metrics (AUC)

Table A1: Performance metrics for SMOTE sample

Grade 3				
Classifier	Literacy		Numeracy	
	AUC	95% CI	AUC	95% CI
Logistic	0.704	(0.699, 0.709)	0.696	(0.691, 0.701)
Elastic Net	0.704	(0.699, 0.709)	0.696	(0.691, 0.701)
Decision Tree	0.617	(0.612, 0.622)	0.614	(0.609, 0.619)
Random Forest	0.697	(0.692, 0.701)	0.689	(0.684, 0.694)
Grade 5+				
Classifier	Literacy		Numeracy	
	AUC	95% CI	AUC	95% CI
Logistic	0.825	(0.823, 0.828)	0.817	(0.815, 0.819)
Elastic Net	0.826	(0.823, 0.828)	0.817	(0.815, 0.820)
Decision Tree	0.771	(0.768, 0.774)	0.764	(0.762, 0.767)
Random Forest	0.824	(0.822, 0.8827)	0.820	(0.818, 0.822)

B Receiver operating characteristic (ROC) curves



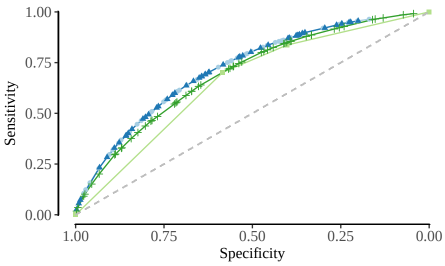
(a) Grade 3 Literacy

(b) Grade 3 Numeracy

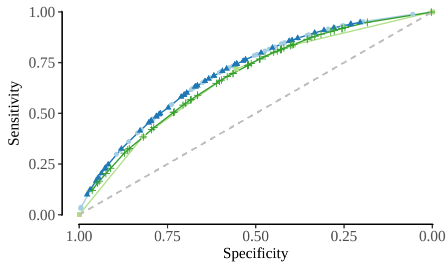
(c) Grade 5+ Literacy

(d) Grade 5+ Numeracy

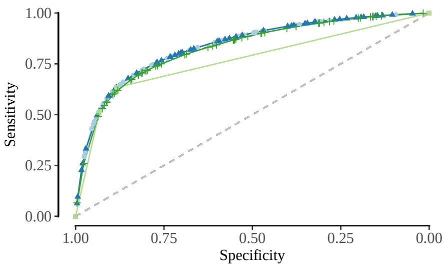
Figure B1: ROC curves for weighted observations



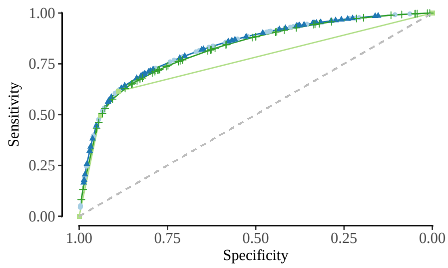
(a) Grade 3 Literacy



(b) Grade 3 Numeracy



(c) Grade 5+ Literacy

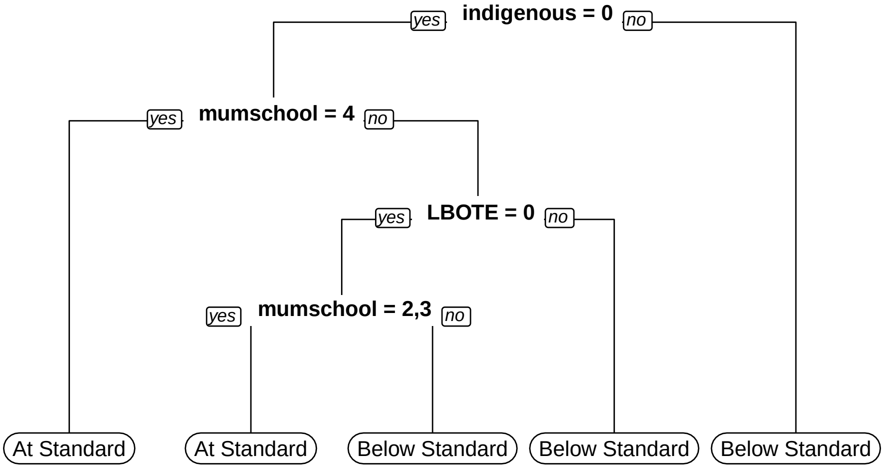


(d) Grade 5+ Numeracy

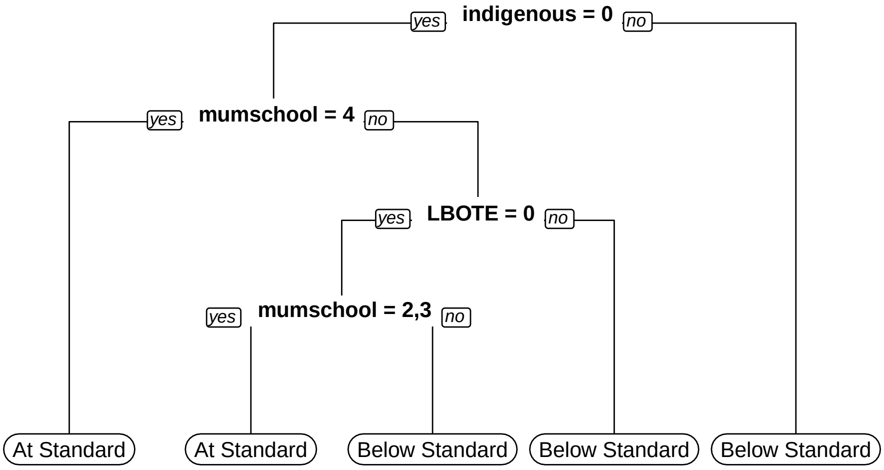
Classifier — Logistic — Elastic Net — Decision Tree — Random Forest

Figure B2: ROC curves for the SMOTE sample

C Decision trees for the SMOTE sample

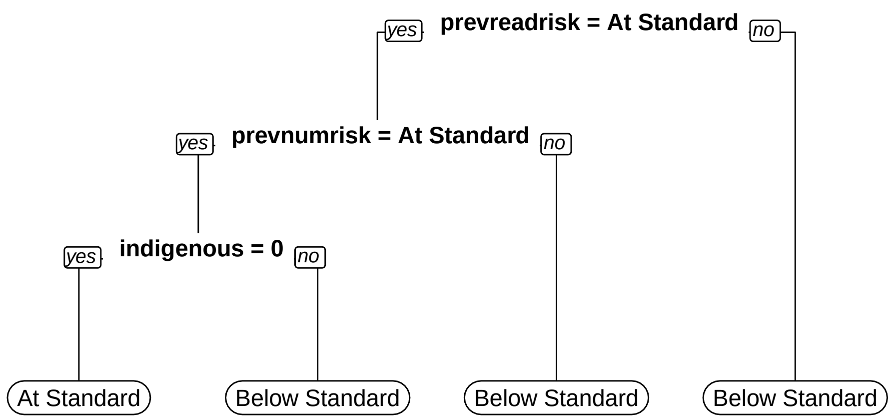


(a) Literacy

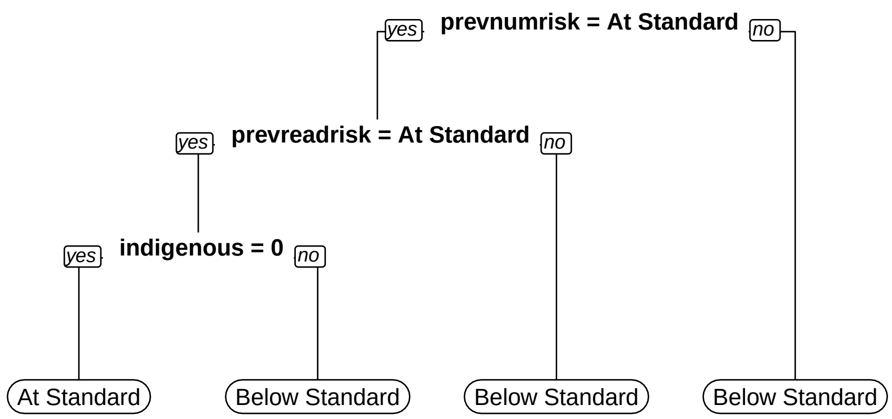


(b) Numeracy

Figure C1: Decison trees for Grade 3 using the SMOTE sample



(a) Literacy



(b) Numeracy

Figure C2: Decison trees for Grade 5+ using the SMOTE sample

CONCLUDING REMARKS

Standardised testing has played an increasingly important role in economic analysis and education policy development. This thesis explored the National Assessment Program - Literacy and Numeracy (NAPLAN) and its uses for both the economist and the policy analyst. While this data set has some particular, and significant, limitations in its ability to achieve causal identification of certain effects on educational outcomes, it remains useful for providing some guidance on the factors determining educational achievement. The goal of understanding how to improve educational outcomes for students may remain difficult, but this thesis situates an Australian research agenda within the international literature, provides some solutions for analysing questions of funding and student background, and develops a future roadmap for the econometric analysis of test scores within the wider framework of big data and modern machine learning techniques.

The first chapter details various testing programmes across a number of countries, before outlining NAPLAN testing in Australia. The goal of this chapter is to provide a description of the uses of this data set, given the types of analysis undertaken using standardised test score data in the international literature. Secondly, we aim to understand the limitations of the NAPLAN data, in order to compare and contrast the kinds of analysis that can be done in Australia and internationally. These limitations are significant, particularly the inability to link the school-level data to the student-level data and the problems associated with following students across time properly. Despite these deficiencies, we define two areas for analysis that may be attempted:

1. Using the school-average level data to analyse school funding in Aus-

tralian schools; and

2. Examining the kinds of students at risk of poor performance on NAPLAN using the student-level test score data.

The next chapters dealt with these research areas in turn.

The second chapter of this thesis attempts to determine the causal effect of per-student school funding on school-average NAPLAN test scores. Using school-average data on test scores, funding, and other covariates, we implement an unconditional panel quantile regression approach, instrumenting with the lagged levels of school funding. Typical OLS estimates may be biased due to reverse causality between test scores and school funding. By instrumenting with funding in the previous period, we account for this reverse causality as the funding in the previous period is exogenous to scores in the current period. The quantile regression approach allows us to understand how funding may impact schools differently across the test score distribution. We find that this is indeed the case, which may explain why the literature has struggled to find a relationship between funding and test scores 'on average'. Specifically, we find that funding is most beneficial when provided at as local a level as possible, indicating the importance of federalist structures for funding and governance of schools. Future analysis of school funding could be improved by nesting the student-level test score data into the school-level data, to understand how changes in funding may affect particular students directly. Alternative future approaches could include difference-in-difference analysis of schools selected into particular funding programs, but this would require further detailed information about funding within these schools. This analysis could also be improved by a better understanding of how principals are allocating their funding to particular budget areas within schools. Nevertheless, this paper reinforces that future policy will benefit from a federalist approach, giving more spending power locally to schools themselves.

In chapter 3, we next turned to the question of student background features and how these interact with performance in both the public and private school systems. This paper uses the student-level test score data, to find what student background features determine NAPLAN achievement. Secondly, we analyse how, given these features, a student's achievement may differ in the public and private schooling sectors. In the course of this analysis, we are faced with the typical sample selection problem into private schools, which is popularly studied in this literature. In order to combat the fact that higher-achieving students may self-select into private schools, we implement a control function approach using mother's employment categories as a candidate instrument for a family's school choice decisions. This variable choice is supported by analysis of the data, which highlights that while a mother's employment category is related to attendance at private school, it does not appear related to test score achievement. Ultimately, we find that private schooling may be useful for some kinds of socially disadvantaged students. This conclusion, however, must be taken with some caveats. Future work may include analysis of the LSAC set of data, to better control for self-selection. Being able to follow students that change schools would also be beneficial for better understanding the effects of school choice on educational achievement. In fact, this paper clearly highlights the difficulties associated with working with the NAPLAN set of data. The lack of background covariates makes it difficult to achieve causal inference, and this paper represents a best attempt to do so. Despite the limitations of this study, it provides an interesting starting point for understanding what is causing poor achievement at the individual-level.

Finally, this thesis explores the predictive capabilities of this data set, given its large sample size and the difficulties involved with causal analysis. While the data remains limited in this sense, it has significant ability for problems of the predictive type. The fourth chapter uses machine learning techniques to illustrate how useful this data can be for the policy-maker, and the school administrator. We show that even with basic background covariates, we are able to predict students that are at risk of poor performance before it occurs. We also find that machine learning techniques are unnecessary to generate good predictions. This paper shows a real, practical example of how this data can be used by the policy-maker to identify poor performance before the fact.

In conclusion, we have learnt much about the NAPLAN data set and the state of schooling in Australia. Yet, there remains much to learn. It is the hope that this thesis can provide a useful starting point for further work in this space, and that it can encourage the economist, the educator, and the policy-maker - at both state and federal levels - to fully commit to working together to improve testing and data collection practices in order to achieve the best evidence base for future policy.

REFERENCES

Abdulkadiroglu, A., Angrist, J. and Pathak, P. (2014), 'The Elite Illusion: Achievement Effects at Boston and New York Exam Schools', *Econometrica* **82**(1), 137–196.

Adams, J. (2018), 'Could machine learning replace the credit score?', *American Banker* **183**(8), 1.

Agrawal, A. and Menzies, T. (2018), Is "Better Data" Better Than "Better Data Miners"?: On the Benefits of Tuning SMOTE for Defect Prediction, *in* 'Proceedings of the 40th International Conference on Software Engineering', ICSE '18, pp. 1050–1061.

Alderman, H., Behrman, J. R., Ross, D. R. and Sabot, R. (1996), 'Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy', *Journal of Human Resources* **31**(1), 229 – 254.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2000), Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, Working Paper 7831, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w7831*

Angrist, J. D. and Krueger, A. B. (1991), 'Does compulsory school attendance affect schooling and earnings?', *Quarterly Journal of Economics* **106**(4), 979 – 1014.

Ashenfelter, O. and Krueger, A. B. (1994), 'Estimates of the Economic Return to Schooling from a New Sample of Twins', *American Economic Review* **84**(5), 1157–1173.

Bacham, D. and Zhao, J. (2017), 'Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling', *Moody's Analytics Risk Perspectives* **9**.

URL: *https://www.moodysanalytics.com/risk-perspectives-magazine/managing-disruption/spotlight/machine-learning-challenges-lessons-and-opportunities-in-credit-risk-modeling*

Baker, M. (2016), 'QREGPD: Stata module to perform Quantile Regression for Panel Data'.

URL: *https://ideas.repec.org/c/boc/bocode/s458157.html*

Baker, M. J. (2014), 'Adaptive Markov chain Monte Carlo sampling and estimation in Mata', *Stata Journal* **14**(3), 623–661.

Ballatore, R. M., Fort, M. and Ichino, A. (2018), 'Tower of babel in the classroom: Immigrants and natives in italian schools', *Journal of Labor Economics* **36**(4).

Baltagi, B. H. and Ghosh, P. K. (2017), 'Replication of Unconditional Quantile Regressions by Firpo, Fortin and Lemieux (2009)', *Journal of Applied Econometrics* **32**(1), 218–223.

Bamber, D. (1975), 'The area above the ordinal dominance graph and the area below the receiver operating characteristic graph', *Journal of Mathematical Psychology* **12**(4), 387 – 415.

Barro, R. J. (1991), 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics* **106**(2), 407–443.

Bast, J. L. and Walberg, H. J. (2004), 'Can parents choose the best schools for their children?', *Economics of Education Review* **23**(4), 431 – 440. Special Issue In Honor of Lewis C. Solman.

Battistin, E., De Nadai, M. and Vuri, D. (2017), 'Counting rotten apples: Student achievement and score manipulation in Italian elementary schools', *Journal of Econometrics* **200**(2), 344–362.

Bedard, K. and Cho, I. (2010), 'Early gender test score gaps across OECD countries', *Economics of Education Review* **29**(3), 348 – 363.

Behrman, J. R. and Knowles, J. C. (1999), 'Household Income and Child Schooling in Vietnam', *World Bank Economic Review* **13**(2), 211 – 256.

Ben-David, A. and Frank, E. (2009), 'Accuracy of machine learning models versus hand crafted expert systems: A credit scoring case study', *Expert Systems with Applications* **36**(3), 5264–5271.

Bitler, M. P., Gelbach, J. B. and Hoynes, H. W. (2006), 'What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments', *American Economic Review* **96**(4), 988–1012.

Bjorklund, A. and Jantti, M. (1997), 'Intergenerational Income Mobility in Sweden Compared to the United States', *American Economic Review* **87**(5), 1009 – 1018.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2011), 'Too Young to Leave the Nest? The Effects of School Starting Age', *Review of Economics and Statistics* **93**(2), 455 – 467.

Blackburn, V., Brennan, S. and Ruggiero, J. (2014), 'Measuring efficiency in

Australian Schools: A preliminary analysis', *Socio-Economic Planning Sciences* **48**(1), 4 – 9.

Bloom, N., Lemos, R., Sadun, R. and Reenen, J. V. (2015), 'Does Management Matter in Schools?', *Economic Journal* **124**(584), 647–674.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**(430), 443–450.

Bourlass, R., Ventelou, B. and Woode, M. E. (2018), 'Child Income Appropriations as a Disease-Coping Mechanism: Consequences for the Health-Education Relationship', *The Journal of Development Studies* **54**(1), 57–71.

Breiman, L. (1996), 'Bagging predictors', *Machine Learning* **24**(2), 123 – 140.

Breiman, L. (2001), 'Random forests', *Machine Learning* **45**(1), 5–32.

Burgess, S., Wilson, D. and Worth, J. (2013), 'A natural experiment in school accountability: The impact of school performance information on pupil progress', *Journal of Public Economics* **106**, 57 – 67.

Canay, I. A. (2011), 'A simple approach to quantile regression for panel data', *The Econometrics Journal* **14**(3), 368–386.

Card, D. (1993), Using geographic variation in college proximity to estimate the return to schooling, NBER Working Papers 4483, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w4483.pdf*

Card, D. (1999), Chapter 30 - The Causal Effect of Education on Earnings, *in* O. C. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Elsevier, pp. 1801 – 1863.

Card, D. and Krueger, A. B. (1994), The Economic Return to School Quality: A Partial Survey, Working Paper 334, Princeton University.
URL: *http://davidcard.berkeley.edu/papers/econ-school.pdf*

Card, D. and Payne, A. (2002), 'School finance reform, the distribution of school spending, and the distribution of student test scores', *Journal of Public Economics* **83**(1), 49 – 82.

Cardak, B. A. and Vecci, J. (2013), 'Catholic school effectiveness in australia: A reassessment using selection on observed and unobserved variables', *Economics of Education Review* **37**, 34 – 45.

Castells-Quintana, D. and Royuela, V. (2017), 'Tracking positive and negative effects of inequality on long-run growth', *Empirical Economics* **53**(4), 1349–1378.

Chakraborty, K. and Harper, R. K. (2017), 'Measuring the impact of socio-economic factors on school efficiency in Australia', *Atlantic Economic Journal* **45**(2).

Chang, Y.-C., Chang, K.-H., Chu, H.-H. and Tong, L.-I. (2016), 'Establishing decision tree-based short-term default credit risk assessment models', *Communications in Statistics - Theory and Methods* **45**(23), 6803–6815.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002), ‘Smote: synthetic minority over-sampling technique’, *Journal of artificial intelligence research* **16**, 321–357.

Chawla, N. V., Lazarevic, A., Halland, L. and Bowyer, K. W. (2003), *Knowledge Discovery in Databases: PKDD 2003. Lecture Notes in Computer Science, vol 2838.*, Springer, Berlin, Heidelberg, chapter SMOTEBoost: Improving Prediction of the Minority Class in Boosting.

Chernozhukov, V. and Hansen, C. (2006), ‘Instrumental quantile regression inference for structural and treatment effect models’, *Journal of Econometrics* **132**(2), 491–525.

Chernozhukov, V. and Hansen, C. (2008), ‘Instrumental variable quantile regression: A robust inference approach’, *Journal of Econometrics* **142**(1), 379 – 398.

Chernozhukov, V. and Hong, H. (2003), ‘An MCMC approach to classical estimation’, *Journal of Econometrics* **115**(2), 293 – 346.

Chetty, R., Friedman, J. N. and Rockoff, J. E. (2014*a*), ‘Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates’, *American Economic Review* **104**(9), 2593–2632.

Chetty, R., Friedman, J. N. and Rockoff, J. E. (2014*b*), ‘Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood’, *American Economic Review* **104**(9), 2633–2679.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014), ‘Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States’, *Quarterly Journal of Economics* **129**(4), 1553 – 1623.

Chinchor, N. and Sundheim, B. (1993), Muc-5 evaluation metrics, *in* 'Proceedings of the 5th conference on Message understanding', Association for Computational Linguistics, pp. 69–78.

Chudgar, A. and Quin, E. (2012), 'Relationship between private schooling and achievement: Results from rural and urban India', *Economics of Education Review* **31**(4), 376 – 390.

Clark, G. (2014), *The Son Also Rises: Surnames and the History of Social Mobility*, Princeton University Press, New Jersey.

Cobb-Clark, D. A. and Jha, N. (2013), Educational Achievement and the Allocation of School Resources, Melbourne Institute Working Paper Series wp2013n27, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.
URL: *https://ideas.repec.org/p/iae/iaewps/wp2013n27.html*

Cobb-Clark, D. A. and Nguyen, T.-H. (2012), 'Educational Attainment Across Generations: The Role of Immigration Background', *Economic Record* **88**(283), 554–575.

Coelli, M., Foster, G. and Leigh, A. (2018), 'Do School Principals Respond to Increased Public Scrutiny? New Survey Evidence from Australia', *Economic Record* **94**, 73 –101.

Collins, C., O'Connor, K. and Yates, L. (2011), *Australia's Curriculum Dilemmas: State Cultures and the Big Issues*, MUP Academic.

Corak, M. (2013), 'Income Inequality, Equality of Opportunity, and Intergenerational Mobility', *Journal of Economic Perspectives* **27**(3), 79 – 102.

Cornell-Farrow, S. (2014), Concerning Learning: A Review of My School and Its Potential Applications to Improving Educational Standards in Developing Countries, Internal document, The Brookings Institution.

Cortez, P. and Silva, A. (2008), 'Using data mining to predict secondary school student performance'.
URL: *http://hdl.handle.net/1822/8024*

Crawford, C., Dearden, L. and Greaves, E. (2014), 'The Drivers of Month-of-Birth Differences in Children's Cognitive and Non-cognitive Skills', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **177**(4), 829 – 860.

Cullen, J. B., Jacob, B. A. and Levitt, S. D. (2005), 'The impact of school choice on student outcomes: an analysis of the Chicago Public Schools', *Journal of Public Economics* **89**(5), 729 – 760.

Cullen, J. B., Levitt, S. D., Robertson, E. and Sadoff, S. (2013), 'What Can Be Done to Improve Struggling High Schools?', *Journal of Economic Perspectives* **27**(2), 133 – 152.

Dancer, D. and Blackburn, V. (2017), 'What are the effective public schools? Insights from New South Wales secondary schools using a Stochastic Frontier Analysis with a panel dataset', *Australian Journal of Education* **61**(2), 141–163.

De Luca, G. and Perotti, V. (2011), 'Estimation of ordered response models with sample selection', *Stata Journal* **11**(2), 213–239(27).

Dearden, L., Ryan, C. and Sibieta, L. (2011), 'What Determines Private School

Choice? A Comparison between the United Kingdom and Australia', *The Australian Economic Review* **44**(3), 308–320.

Deller, S. C. and Rudnicki, E. (1993), 'Production efficiency in elementary education: The case of Maine public schools', *Economics of Education Review* **12**(1), 45 – 57.

Dewey, J., Husted, T. A. and Kenny, L. W. (2000), 'The Ineffectiveness of School Inputs: A Product of Misspecification?', *Economics of Education Review* **19**(1), 27–45.

Di Porto, E., Parenti, A., Paty, S. and Abidi, Z. (2017), 'Local Government Cooperation at Work: A Control Function Approach', *Journal of Economic Geography* **17**(2), 435 – 463.

Dickerson, A., McIntosh, S. and Valente, C. (2015), 'Do the maths: An analysis of the gender gap in mathematics in Africa', *Economics of Education Review* **46**, 1–22.

Ding, W. and Lehrer, S. F. (2014), 'Understanding the role of time-varying unobserved ability heterogeneity in education production', *Economics of Education Review* **40**, 55 – 75.

Doan, B. (2016), 'Responsive affirmative action in school choice', *Journal of Economic Theory* **165**, 69 – 105.

Einav, L. and Levin, J. (2014), 'Economics in the age of big data', *Science* **346**(6210).

Elakia, G. and Aarthi, N. J. (2014), 'Application of data mining in educational database for predicting behavioural patterns of the students', *International Journal of Computer Science and Information Technologies* **5**(3), 4649–4652.

Elder, T. and Jepsen, C. (2014), 'Are Catholic primary schools more effective than public primary schools?', *Journal of Urban Economics* **80**, 28–38.

Elkan, C. (2001), The foundations of cost-sensitive learning, *in* 'International joint conference on artificial intelligence', Vol. 17, pp. 973–978.

Ellison, G. and Swanson, A. (2016), 'Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions', **106**(6), 1244 – 1277.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), 'Unconditional Quantile Regressions', *Econometrica* **77**(3), 953–973.

Fletcher, J. and Kim, T. (2016), 'The effects of changes in kindergarten entry age policies on educational achievement', *Economics of Education Review* **50**, 45 – 62.

Ford, M. (2013), 'Achievement Gaps in Australia: What NAPLAN reveals about Education Inequality in Australia', *Race Ethnicity and Education* **16**(1), 80–102.

Fredriksson, P. and Ockert, B. (2014), 'Life-Cycle Effects of Age at School Start', *Economic Journal* **124**(579), 977 – 1004.

Friedman, J., Hastie, T. and Tibshirani, R. (2001), *The elements of statistical learning*, Vol. 1, Springer series in statistics New York, NY, USA:.

Friedman, J., Hastie, T. and Tibshirani, R. (2009), 'glmnet: Lasso and elastic-net regularized generalized linear models', *R package version* **1**(4).
URL: *https://cran.r-project.org/package=glmnet*

Frolich, M. and Melly, B. (2013), 'Unconditional Quantile Treatment Effects under Endogeneity', *Journal of Business and Economic Statistics* **31**(3), 346 – 357.

Fryer, Roland G., J. and Levitt, S. D. (2004), 'Understanding the Black-White Test Score Gap in the First Two Years of School', *Review of Economics and Statistics* **86**(2), 447 – 464.

Fryer, Roland G., J. and Levitt, S. D. (2005), The Black-White Test Score Gap Through Third Grade, NBER Working Papers 11049, National Bureau of Economic Research.

Fryer, Roland G., J. and Levitt, S. D. (2010), 'An Empirical Analysis of the Gender Gap in Mathematics', *American Economic Journal: Applied Economics* **2**(2), 210 – 240.

Fryer, Roland G., J., Levitt, S. D. and List, J. A. (2015), Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights, NBER Working Papers 21477, National Bureau of Economic Research.

Galvao Jr, A. F. (2011), 'Quantile regression for dynamic panel data with fixed effects', *Journal of Econometrics* **164**(1), 142–157.

Gibbons, S. and Silva, O. (2011), 'Faith Primary Schools: Better Schools or Better Pupils?', *Journal of Labor Economics* **29**(3), 589–635.

Gibbons, S., Silva, O. and Weinhardt, F. (2013), 'Everybody needs good neighbours? Evidence from students' outcomes in England', *The Economic Journal* **123**(571), 831–874.

Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., Perry, L.-A. and Roberts, M. (2018), *Through Growth to Achievement: Report of the Review to Achieve Educational Excellence in Australian Schools*, Department of Education, Employment and Workplace Relations, Canberra.
URL: *docs.education.gov.au/*

Gonski, D., Boston, K., Greiner, K., Lawrence, C., Scales, B. and Tannock, P. (2011), *Review of Funding for Schooling - Final Report*, Department of Education, Employment and Workplace Relations, Canberra.
URL: *docs.education.gov.au/*

Goss, P. and Sonnemann, J. (2016), Widening gaps: What NAPLAN tells us about student progress, Report, Grattan Institute.
URL: *https://grattan.edu.au/report/widening-gaps/*

Goss, P. and Sonnemann, J. (2018), Measuring student progress: A state-by-state report card, Report, Grattan Institute.
URL: *https://grattan.edu.au/report/measuring-student-progress*

Graddy, K. and Stevens, M. (2005), 'The Impact of School Resources on Student Performance: A Study of Private Schools in the United Kingdom', *Industrial & Labor Relations Review* **58**(3), 435–451.

Gray, G., McGuinness, C. and Owende, P. (2013), An investigation of psychometric measures for modelling academic performance in tertiary education, *in* 'Educational Data Mining 2013'.

Green, F., Machin, S., Murphy, R. and Zhu, Y. (2011), 'The Changing Economic Advantage From Private Schools', *Economica* **79**(316), 658–679.

Guegan, D. and Hassani, B. (2018), 'Regulatory learning: How to supervise machine learning models? An application to credit scoring', *The Journal of Finance and Data Science* **4**(3), 157 – 171.

Hahn, M. H. and Haisken-DeNew, J. P. (2013), 'PanelWhiz and the Australian Longitudinal Data Infrastructure in Economics', *Australian Economic Review* **46**(3), 379 – 386.

Hamilton, L. S., Stecher, B. M. and Klein, S. P. (2002), *Making Sense of Test-Based Accountability in Education*, RAND Corporation, Santa Monica.

Hanushek, E. (2011), 'The economic value of higher teacher quality', *Economics of Education Review* **30**(3), 466–479.

Hanushek, E. A. (1979), 'Conceptual and empirical issues in the estimation of educational production functions', *Journal of Human Resources* **14**(3), 351–388.

Hanushek, E. A. (2003), 'The failure of input-based schooling policies', *Economic Journal* **113**(485), 64–98.

Hanushek, E. A. (2013), 'Economic growth in developing countries: The role of human capital', *Economics of Education Review* **37**, 204–212.

Hanushek, E. A., Kain, J. F., Rivkin, S. G. and Branch, G. F. (2007), 'Charter school quality and parental decision making with school choice', *Journal of Public Economics* **91**(5–6), 823 – 848.

Hanushek, E. A. and Raymond, M. F. (2005), 'Does School Accountability Lead to Improved Student Performance?', *Journal of Policy Analysis & Management* **24**(2), 297 – 327.

Hanushek, E. A. and Rivkin, S. G. (2010), 'Generalizations about Using Value-Added Measures of Teacher Quality', *American Economic Review* **100**(2), 267 – 271.

Hanushek, E. A., Rivkin, S. G. and Taylor, L. L. (1996), Aggregation and the Estimated Effects of School Resources, NBER Working Papers 5548, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w5548.pdf*

Hanushek, E. A. and Woessman, L. (2008), 'The role of cognitive skills in economic development', *Journal of Economic Literature* **46**(3), 607–668.

Hanushek, E. A. and Woessmann, L. (2011), 'Sample selectivity and the validity of international student achievement tests in economic research', *Economics Letters* **110**(2), 79 – 82.

Harding, M. and Lamarche, C. (2009), 'A quantile regression approach for estimating panel data models using instrumental variables', *Economics Letters* **104**(3), 133–135.

Harris, D. (2010), Education production functions: Concepts, *in* D. J. Brewer and P. J. McEwan, eds, 'Economics of Education', Academic Press, Oxford.

Hausman, J. (1978), 'Specification Tests in Econometrics', *Econometrica* **46**(6), 1251–1271.

Heckman, J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**, 153–161.

Hoerl, A. E. and Kennard, R. W. (1970), 'Ridge Regression: Biased Estimation for Nonorthogonal Problems', *Technometrics* **12**(1), 55–67.

Hsieh, C.-T. and Urquiola, M. (2006), 'The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program', *Journal of Public Economics* **90**(8), 1477 – 1503.

Huang, S. and Fang, N. (2013), 'Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models', *Computers & Education* **61**, 133 – 145.

Jacob, R. T., Goddard, R. D. and Kim, E. S. (2014), 'Assessing the Use of Aggregate Data in the Evaluation of School-Based Interventions: Implications for Evaluation Research and State Policy Regarding Public-Use Data', *Educational Evaluation and Policy Analysis* **36**(1), 44–66.

Japkowicz, N. et al. (2000), Learning from imbalanced data sets: a comparison of various strategies, *in* 'AAAI workshop on learning from imbalanced data sets', Vol. 68, Menlo Park, CA, pp. 10–15.

Jarner, S. F. and Roberts, G. O. (2007), 'Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms', *Scandinavian Journal of Statistics* **34**(4), 781 – 815.

Jensen, P. and Rasmussen, A. W. (2011), 'The effect of immigrant concentration in schools on native and immigrant children's reading and math skills', *Economics of Education Review* **30**(6), 1503 – 1515.

Jishan, S. T., Rashu, R. I., Haque, N. and Rahman, R. M. (2015), 'Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique', *Decision Analytics* **2**(1), 1.

Kalb, G. and van Ours, J. C. (2014), 'Reading to young children: A head-start in life?', *Economics of Education Review* **40**, 1 – 24.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z. (2015), 'Prediction Policy Problems', *American Economic Review* **105**(5), 491 – 495.

Knight, K. and Fu, W. (2000), 'Asymptotics for lasso-type estimators', *The Annals of Statistics* **5**(28), 1356–1378.

Koenker, R. (2004), 'Quantile regression for longitudinal data', *Journal of Multivariate Analysis* **91**(1), 74–89.

Kotsiantis, S. B. (2012), 'Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades', *Artificial Intelligence Review* **37**(4), 331–344.

Kotsiantis, S., Pierrakeas, C. and Pintelas, P. (2004), 'Predicting student's performance in distance learning using machine learning techniques', *Applied Artificial Intelligence* **18**(5), 411–426.

Kovacic, Z. (2010), 'Early prediction of student success: Mining students' enrolment data'.

URL: *http://hdl.handle.net/11072/646*

Lamarche, C. (2010), 'Robust penalized quantile regression estimation for panel data', *Journal of Econometrics* **157**(2), 396–408.

Lazear, E. P. (2006), 'Speeding, Terrorism, and Teaching to the Test', *Quarterly Journal of Economics* **121**(3), 1029 – 1061.

Le Van, C., Nguyen, A. N., Nguyen, N.-M. and Simioni, M. (2018), 'Growth strategy with social capital, human capital and physical capital?Theory and evidence: The case of Vietnam', *Journal of Public Economic Theory* . Corrected Proof.

Lee, L.-F., Maddala, G. S. and Trost, R. P. (1980), 'Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity', *Econometrica* **48**(2), 491–503.

Lefebvre, P., Merrigan, P. and Verstraete, M. (2011), 'Public subsidies to private schools do make a difference for achievement in mathematics: Longitudinal evidence from Canada', *Economics of Education Review* **30**(1), 79–98.

Lemieux, T. and Card, D. (2001), 'Education, Earnings, and the 'Canadian G. I. Bill'', *Canadian Journal of Economics* **34**(2), 313 – 344.

Levin, H. M. (1974), 'Measuring efficiency in educational production', *Public Finance Review* **2**(1), 3–24.

Levin, H. M. (1998), 'Educational vouchers: Effectiveness, choice and costs', *Journal of Policy Analysis and Management* **17**(3), 373–392.

Levitt, S. D., List, J. A., Neckerman, S. and Sadoff, S. (2016), The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance, NBER Working Papers 18165, National Bureau of Economic Research.

Liu, X., Lovely, M. E. and Ondrich, J. (2010), 'The Location Decisions of Foreign Investors in China: Untangling the Effect of Wages Using a Control Function Approach', *Review of Economics and Statistics* **92**(1), 160 – 166.

Luo, C., Wu, D. and Wu, D. (2017), 'A deep learning approach for credit scoring using credit default swaps', *Engineering Applications of Artificial Intelligence* **65**, 465 – 470.

McDermott, K. (2011), *High-Stakes Reform: The Politics of Educational Accountability*, Georgetown University Press, Baltimore.

McVicar, D., Moschion, J. and Ryan, C. (2018), 'Achievement effects from new peers: Who matters to whom?', *Economics of Education Review* **66**, 154 – 166.

Mendolia, S. and Siminski, P. (2017), 'Is education the mechanism through which family background affects economic outcomes? a generalised approach to mediation analysis', *Economics of Education Review* **59**, 1 – 12.

Mengersen, K. L. and Tweedie, R. L. (1996), 'Rates of convergence of the Hastings and Metropolis algorithms', *The Annals of Statistics* **24**(1), 101–121.

Miller, P. W. and Voon, D. (2011), 'Lessons from My School', *Australian Economic Review* **44**(4), 366–386.

Miller, P. W. and Voon, D. (2012), 'Government Versus Non-Government Schools: A Nation-Wide Assessment Using Australian NAPLAN Data', *Australian Economic Papers* **51**(3), 147–166.

Miller, P. W. and Voon, D. (2014), 'School outcomes in New South Wales and Queensland: a regression discontinuity approach', *Education Economics* **22**(5), 427–448.

Mishra, T., Kumar, D. and Gupta, S. (2014), Mining Students' Data for Prediction Performance, *in* '2014 Fourth International Conference on Advanced Computing Communication Technologies', pp. 255–262.

Morris, A. (2011), Student Standardised Testing: Current Practices in OECD Countries and a Literature Review, OECD Education Working Papers 65, OECD Publishing.
URL: *http://dx.doi.org/10.1787/5kg3rp9qbnr6-en*

Mullainathan, S. and Spiess, J. (2017), 'Machine Learning: An Applied Econometric Approach', *Journal of Economic Perspectives* **31**(2), 87–106.

Murtazashvili, I. and Wooldridge, J. M. (2016), 'A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching', *Journal of Econometrics* **190**(2), 252 – 266.

Natek, S. and Zwilling, M. (2014), 'Student data mining solution-knowledge management system related to higher education institutions', *Expert Systems with Applications* **41**(14), 6400 – 6407.

National Commission of Audit (2014), 'Towards Responsible Government: Appendix to the Report of the National Commission of Audit - Volume 1: 9.7 Schools funding.'.
URL: *http://www.ncoa.gov.au/report/*

Nghiem, H. S., Nguyen, H. T., Khanam, R. and Connelly, L. B. (2015), 'Does school type affect cognitive and non-cognitive development in children? Evidence from Australian primary schools', *Labour Economics* **33**, 55–65.

Nghiem, S., Nguyen, H. T. and Connelly, L. B. (2016), 'The Efficiency of Australian Schools: A Nationwide Analysis Using Gains in Test Scores of Students as Outputs', *Economic Papers* **35**(3), 256 – 268.

Nicoletti, C. and Rabe, B. (2013), 'Inequality in Pupils' Test Scores: How Much Do Family, Sibling Type and Neighbourhood Matter?', *Economica* **80**(318), 197 – 218.

Pandey, P., Goyal, S. and Sundararaman, V. (2009), 'Community Participation in Public Schools: Impact of Information Campaigns in Three Indian States', *Education Economics* **17**(3), 355 – 375.

Parish, R. (1963), 'The Economics of State Aid to Education', *Economic Record* **39**(87), 292–304.

Peterson, P. E. and West, M. R. (2003), *No Child Left Behind? The Politics and Practice of School Accountability*, The Brookings Institution, Washington DC.

Polidano, C., Hanel, B. and Buddelmeyer, H. (2013), 'Explaining the Socioeconomic Status School Completion Gap', *Education Economics* **21**(3), 230 – 247.

Polidano, C. and Jha, N. (2015), 'Long-Run Effects of Catholic Schooling on Wages', *The B.E. Journal of Economic Analysis & Policy* **15**(4), 2017–2045.

Polidano, C. and Ryan, C. (2017), 'What Happens to Students with Low

Reading Proficiency at 15? Evidence from Australia', *Economic Record* **93**(303), 600–614.

Powell, D. (2014), Did the Economic Stimulus Payments of 2008 Reduce Labor Supply? Evidence from Quantile Panel Data Estimation, Working Paper 710-3, RAND Corporation.

Powell, D. (2015), Quantile Regression with Nonadditive Fixed Effects, Selected Works of David Powell, RAND Corporation.

Powell, D. (2016), Quantile Treatment Effects in the Presence of Covariates, Selected Works of David Powell, RAND Corporation.

Pugh, K. and Foster, G. (2014), 'Australia's National School Data and the 'Big Data' Revolution in Education Economics', *The Australian Economic Review* **47**(2), 258–268.

Reback, R. (2008), 'Teaching to the rating: School accountability and the distribution of student achievement', *Journal of Public Economics* **92**(5), 1394 – 1415.

Roberts, G., Gelman, A. and Gilks, W. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *The Annals of Applied Probability* **7**(1), 110–120.

Roberts, G. O. and Rosenthal, J. S. (2001), 'Optimal scaling for various Metropolis-Hastings algorithms', *Statistical Science* **16**(4), 351–367.

Roberts, G. O. and Rosenthal, J. S. (2009), 'Examples of Adaptive MCMC', *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Romero, C. and Ventura, S. (2007), 'Educational data mining: A survey from 1995 to 2005', *Expert Systems with Applications* **33**(1), 135 – 146.

Ryan, C. (2013), 'What is behind the decline in student achievement in australia?', *Economics of Education Review* **37**, 226 – 239.

Ryan, C. (2017*a*), 'Measurement of Peer Effects', *Australian Economic Review* **50**(1), 121 – 129.

Ryan, C. (2017*b*), Secondary School Teacher Effects on Student Achievement in Australian Schools, Melbourne institute working paper series, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

URL: *https://ideas.repec.org/p/iae/iaewps/wp2017n11.html*

Santin, D. and Sicilia, G. (2015), 'Measuring the efficiency of public schools in Uruguay: Main drivers and policy implications', *Latin American Economic Review* **24**, 1–28.

Semykina, A. (2017), 'Self-employment among women: Do children matter more than we previously thought?', *Journal of Applied Econometrics* **33**(3), 416–434.

Semykina, A. and Wooldridge, J. M. (n.d.), 'Binary response panel data models with sample selection and self-selection', *Journal of Applied Econometrics* **33**(2), 179–197.

Shafiq, M. N. (2013), 'Gender Gaps in Mathematics, Science and Reading Achievements in Muslim Countries: A Quantile Regression Approach', *Education Economics* **21**(4), 343 – 359.

Shingari, I., Kumar, D. and Khetan, M. (2017), 'A review of applications of data mining techniques for prediction of students performance in higher education', *Journal of Statistics and Management Systems* **20**(4), 713–722.

Stevens, A. H. and Schaller, J. (2011), 'Short-run effects of parental job loss on children's academic achievement', *Economics of Education Review* **30**(2), 289 – 299.

Summers, A. A. and Wolfe, B. L. (1977), 'Do Schools Make a Difference?', *American Economic Review* **67**(4), 639 – 652.

Suziedelyte, A. and Zhu, A. (2015), 'Does early schooling narrow outcome gaps for advantaged and disadvantaged children?', *Economics of Education Review* **45**, 76 – 88.

Tavares, P. A. (2015), 'The impact of school management practices on educational performance: Evidence from public schools in Sao Paulo', *Economics of Education Review* **48**, 1–15.

Terza, J. V., Basu, A. and Rathouz, P. J. (2008), 'Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling', *Journal of Health Economics* **27**(3), 531–543.

Therneau, T., Atkinson, B. and Ripley, B. (2015), 'rpart: Recursive Partitioning and Regression Trees. R package version 4.1–10'.
URL: *https://cran.r-project.org/package=rpart*

Tibshirani, R. (1996), 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.

Todd, P. E. and Wolpin, K. I. (2003), 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *The Economic Journal* **113**(485), F3 – F33.

Torgo, L. and Torgo, M. L. (2013), 'Package DMwR', *Comprehensive R Archive Network* .
URL: *https://cran.r-project.org/package=DMwR*

Vandamme, J., Meskens, N. and Superby, J. (2007), 'Predicting Academic Performance by Data Mining Methods', *Education Economics* **15**(4), 405–419.

Varian, H. R. (2014), 'Big Data: New Tricks for Econometrics', *Journal of Economic Perspectives* **28**(2), 3–28.

Victorian Department of Education and Training (2015), 'Parent Payments in Government Schools'.
URL: *http://www.education.vic.gov.au*

Wan, L., Zeiler, M., Zhang, S., Cun, Y. L. and Fergus, R. (2013), Regularization of Neural Networks using DropConnect, *in* S. Dasgupta and D. McAllester, eds, 'Proceedings of the 30th International Conference on Machine Learning', Vol. 28.3 of *Proceedings of Machine Learning Research*, PMLR, Atlanta, Georgia, USA, pp. 1058–1066.
URL: *http://proceedings.mlr.press/v28/wan13.html*

Warren, D. and Haisken-DeNew, J. P. (2013), Early Bird Catches the Worm: The Causal Impact of Pre-school Participation and Teacher Qualifications on Year 3 National NAPLAN Cognitive Tests, Melbourne institute working paper series, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Watson, L. and Ryan, C. (2010), 'Choosers and losers: The impact of government subsidies on Australian secondary schools', *Australian Journal of Education (ACER Press)* **54**(1), 86 – 107.

Williams, R. A. (1985), 'The Economic Determinants of Private Schooling in Australia', *Economic Record* **61**(174), 622.

Woessman, L., Ludemann, E., Schutz, G. and West, M. R. (2007), School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003, OECD Education Working Papers 13, OECD Publishing.
URL: *https://ideas.repec.org/p/oec/eduaab/13-en.html*

Wooldridge, J. M. (2015), 'Control Function Methods in Applied Econometrics', *Journal of Human Resources* **50**(2), 420 – 445.

Wright, M. N., Wager, S. and Probst, P. (2018), 'ranger: A Fast Implementation of Random Forests'.
URL: *https://cran.r-project.org/package=ranger*

Yadav, S. K. and Pal, S. (2012), 'Data mining: A prediction for performance improvement of engineering students using classification', *arXiv preprint arXiv:1203.3832* .

Yeung, T. Y.-C. (2017), 'Political philosophy, executive constraint and electoral rules', *Journal of Comparative Economics* **45**(1), 67 – 88. Institutions and Economic Change.

Zimmer, R., Gill, B., Booker, K., Lavertu, S. and Witte, J. (2012), 'Examining charter student achievement effects across seven states', *Economics of Education Review* **31**(2), 213 – 224.

Zou, H. and Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320.