The Efficiency of Australian Schools: Evidence from the NAPLAN Data 2009-2011

Nghiem, Son and Nguyen, Ha and Connelly, Luke

The University of Queensland

27 May 2014

The Efficiency of Australian Schools: Evidence from the NAPLAN Data 2009-2011 *

Hong Son Nghiem[a], Ha Trong Nguyen[a] and Luke B. Connelly[a,b]

[a] Centre of National Research on Disability and Rehabilitation Medicine, The University of Queensland

[b] Australian Centre for Economic Research on Health, School of Economics and Centre for Clinical Research,The University of Queensland

Abstract

This study examines the technical efficiency of schools in Australia and its determinants using NAPLAN test results of about 6,800 schools in 2009-2011 and other information from the "My School" website. For each school, we use the average growth of test scores for the same students between 2009 and 2011 as the measure of the school's output and four input measures: the student-teacher ratios, student-non-teaching staff ratios, recurrent income per student and (averaged) capital expenditure per student. We are also able to compare schools by type: including whether or not the school is a public school or a private school, a single sex or co-educational schools, a primary or secondary school, or a school that provides both primary and secondary schooling. In addition we control for several other environmental indicators for each school including: an index of social and educational advantage, the proportion of school children who identify as an Aborigine or Torres Strait Islander, the proportion of students from a non English-speaking background, the proportion of students female, as well as the region, state and territory in which the school is located. We estimate that the average technical efficiency score of Australian schools is 59 per cent and find evidence of input congestion for all of the inputs studied. On average, the growth target for schools in the sample to reach the efficiency frontier is 100 NAPLAN points. Our results suggest that eliminating inputs congestion could, in theory, reduce expenditure per school student by A$2,000. At the primary level, Catholic and independent schools are less efficient than public schools, but this story is reversed at the secondary level. We also find that schools with students from more advantageous social and economic backgrounds and schools with higher ratios of students from non-English speaking backgrounds tend to be more efficient. The results are robust to the choices about how to construct the frontier (e.g., in aggregate or for disaggregates by school type) and to our treatments of output and super-efficiency.

Key words: Efficiency, Australia, data envelopment analysis, double bootstrap

JEL classification: I21, D24

*The data used in this publication are sourced from the Australian Curriculum, Assessment and Reporting Authority (ACARA) and are available from ACARA in accordance with its Data Access Protocols.

1 Introduction

The productivity and efficiency of education has long been of interest to policy-makers, educators and parents worldwide. In Australia, the debate about school efficiency and productivity has intensified in recent years. In 2008, all Australian Education Ministers released the Melbourne Declaration on Educational Goals for Young Australians, setting out the future directions for Australian schooling in the next 10 years (MCEERYA, 2008b). To support the Melbourne Declaration, a series of action plans including curriculum designs, school assessments and financing have been proposed and implemented (MCEERYA, 2008a). One of the major reforms that was instituted was the introduction of a National Assessment Program—Literacy and Numeracy (NAPLAN) in 2008. The NAPLAN was initiated to provide a "vigorous and comprehensive" assessment of student progress across Australia. To strengthen accountability and transparency of schooling, results of NAPLAN tests are then made available to the public via a website called "MySchool".[1]

The availability of school average test scores on the MySchool website as well as a range of other indicators of schools' characteristics provides a new opportunity to compare the performance of schools across Australia. In this paper, we take advantage of the availability of the NAPLAN results to provide robust estimates of school performance. We do so by examining the technical efficiency of almost all Australian schools and by investigating the variations in input combinations and environmental factors that affect the level of efficiency achieved by schools. As far as we are aware, Haug and Blackburn (2013a) is only existing study that uses NAPLAN test scores to examine the efficiency of public secondary schools, and it does so for the State of New South Wales (NSW).

Our study extends previous work along several dimensions. First, rather than focusing on schools in one state, we examine the technical efficiency of virtually all mainstream schools in Australia.[2] Second, unlike previous work that has focused on public schools, this study focuses both on public and non-public schools and also distinguishes between Catholic and types of independent schools. This enables comparisons of the relative performance of schools from different sectors and schools of different types (e.g., single-sex and co-educational schools). Third, we also investigate efficiency of primary schools. By providing robust estimates of school performance on almost all mainstream Australian schools, this study thus extends the literature on school efficiency. The results are likely to be of interest to a broad group of interests, from consumers, to educators and policy-makers (Gonski et al., 2011; Masters, 2012).

Using the NAPLAN test score gains of the same students between 2009 and 2011 as a measure of school output and a double bootstrapping Data Envelopment Analysis (DEA) method, we find that the average technical efficiency of Australian Schools is 59 percent. At the primary school level, Catholic and in-

[1] MySchool website (www.myschool.edu.au) was launched for the first time in 2010.

[2] We exclude special schools that educate children with disabilities and distance education schools from the study for reasons that are provided below.

dependent schools are less efficient than public schools. At the secondary school level, though, public schools are found to be less efficient than other non-public schools. By jurisdiction, both the primary and secondary school levels, schools in NSW and the ACT are estimated to be the most efficient. We also find that schools that draw students from more advantageous social-educational backgrounds as well as schools that have higher ratios of students from non-English speaking backgrounds tend to be more efficient.

The remaining of this paper is structured as follows. Section 2 provides a brief review of literature on school performance. Section 3 presents the method used to investigate school efficiency and its determinants. Section 4 describes the data. Section 5 presents our empirical results and Section 6 concludes the paper.

2 Literature review

The literature on school performance is vast, so our review focuses strictly on studies that are relevant to efficiency measurement in the school sector. The review is partitioned into two components, the first of which reviews the international literature and the second of which reviews the Australian literature. Our discussion of the international literature focuses on the variety of methodological approaches that have been used, while our discussion of the Australian literature concerns the methods, data and findings of the small number of studies on school efficiency for that country.

2.1 International studies

There is a large international literature on school efficiency and productivity and it incorporates a range of different approaches to the measurement of these concepts. For instance, a number of efficiency studies assume that schools are output maximizers and adopt an output-based approach (Bradley et al., 2001; Grosskopf and Moutray, 2001; Grosskopf et al., 2009) whereas other studies start from the (dual) assumption that the school's objective is to minimize the costs of the inputs it uses to achieve its output (Gronberg et al., 2012). The output-based literature also uses a variety of output measures, including the number of students (Ouellette and Vierstraete, 2005; Burney et al., 2011), measures of students' academic performance (Bradley et al., 2001; Alexander et al., 2010; Agasisti, 2011; Kirjavainen, 2011; Mancebón et al., 2012), measures of students' non-academic performance (Bradley et al., 2001) changes in students' academic performance (Grosskopf et al., 2009; Gronberg et al., 2012), or combinations of the foregoing outcome measures (Bradley et al., 2001; Gronberg et al., 2012). By contrast, studies that follow an input approach typically measure inputs by adopting a monetized value, namely the cost per student (Chakraborty and Poggio, 2008; Gronberg et al., 2012; Haelermans et al., 2012).

Another source of variation in the approach to productivity and efficiency measurement in the international literature is the choice of technique that is applied. Some of the existing literature (Chakraborty and Poggio, 2008; Conroy and Arguea, 2008; Kirjavainen, 2011; Gronberg et al., 2012) applies parametric methods such as stochastic frontier analysis (SFA), while others apply non-parametric methods such as Data Envelopment Analysis (DEA): Grosskopf and Moutray (2001); Grosskopf et al. (2009); Alexander et al. (2010); Agasisti (2011); Essid et al. (2011); Haelermans and Ruggiero (2013).

Finally, the international literature on school performance also differs in the educational levels that are examined: some studies focus on primary schooling (Conroy and Arguea, 2008), while others focus on secondary schooling (Alexander et al., 2010; Mancebón et al., 2012; Grosskopf and Moutray, 2001; Kirjavainen, 2011; Haug and Blackburn, 2013a), mixed level education (Burney et al., 2011; Blackburn et al., 2013), or tertiary education (Zoghbi et al., 2013). Some studies have also attempted to compare relative performance of across different education sectors, comparing public and private schools (Cherchye et al., 2010; Mancebón et al., 2012), or charter schools and traditional public schools (Grosskopf et al., 2009; Gronberg et al., 2012).

2.2 Australian studies

A number of studies have attempted to assess the performance of Australian schools using either state-level or national data. As far as we know, the study by Mante and O'Brien (2002) was the first Australian study on school efficiency to use the DEA approach. The authors examined the technical efficiency of 27 public secondary schools in the State of Victoria for the year 1996, using the proportion of students with tertiary entrance scores of 50 and above and the Year 12 apparent retention rates as output measures and the number of staff per student and expenditure per student as input measures. The authors found that although most Victorian schools were performing at a fairly high level, relative to each other, but that most schools were in a position to improve their output. The authors noted the importance of controlling for the socio-economic status of students in their study: omitting this variable would have resulted in some schools being identified as inefficient when, in fact, lower performance was due to intakes of students from lower socio-economic backgrounds.

Lamb et al. (2004) also examined patterns of schools performance in Victoria in the early 2000s. The study of Lamb et al. (2004) differs from that of Mante and O'Brien (2002) in several ways. First, Lamb et al. (2004) studied both primary and secondary schools. Second, they measured school efficiency by calculating the standardized residuals from regressions of school academic outcomes on a set of variables that describe school characteristics such as prior academic achievement but excludes measures of resources such as labor and expenditure. Lamb et al. (2004) found that the socio-economic backgrounds of students exerted a major influence over educational outcomes, that students were highly segregated along social and academic lines, noting that this segregation and sorting intensifies the differences in outcomes between students. Notwithstanding these observations, the authors found examples of efficient schools in each sector, but also found considerable variation in school efficiency for both primary and secondary schools. Their results also indicated that, at the secondary school level, independent schools were the most efficient at improving students' academic outcomes, followed by Catholic schools, with public schools being the least efficient.[3] School funding (as measured either by total income per student for public schools or tuition fees per student for independent schools) was found to be positively associated with school efficiency

Bradley et al. (2004) focused on the performance of public primary schools in the State of Queensland in 2001. They used both output-based and input-based approaches, taking the average of scaled numeracy and literacy scores for each school at Year 7 as the outcome variable and applying DEA. Their results also highlight the importance of controlling for the socio-economic backgrounds of children and for the quality of student intake when assessing school efficiency. Based on greater efficiency scores in local government areas with greater school density, they suggest that increasing competition between government schools may hold promise as a way to increase school efficiency.

More recently, three further studies have used data from Australia's most populous state, New South Wales (NSW), to examine the performance of public schools there Chakraborty and Blackburn (2013); Haug and Blackburn (2013b,a). Blackburn et al. (2013) examined the efficiency of NSW public primary and secondary schools in 2010. Using an output-based approach, the authors examine test scores at the third- and fifth-grades for primary schools and the seventh- and ninth-grades for secondary schools– as outputs in a DEA model. They also used log of per student expenditure as the dependent variable in a cost function approach and the inefficiency-effects model proposed by Battese and Coelli (1995) and obtain similar results. Blackburn et al. (2013) found that schools in NSW are moderately inefficient and that schools with more favorable environments–as measured by the socio-economic background of students and school location–and schools with larger enrolments tend to be more efficient. Similarly, Chakraborty and Blackburn (2013) measured the cost efficiency of public primary and secondary schools for the period 2008-2010. They use the same outputs as were used by Blackburn et al. (2013) but apply a more complex model that contains a richer set of controls for school resources and implementing a two-stage DEA approach. Chakraborty and Blackburn (2013) find that, on average, primary schools are 88 per cent cost-efficient and that secondary schools are 89 per cent cost-efficient. They also find that social disadvantage is strongly negatively associated with the efficiency of primary schools (only). They also find that, over the study period, primary schools' cost-efficiency has generally decreased, while the cost-efficiency of secondary schools increased slightly.

Haug and Blackburn (2013a) also studied the efficiency of public secondary schools in NSW during 2008-2010. The authors used three value-added (i.e., achievement growth) measures of academic results as outcomes: (i) the difference between schools' median Year 12 Higher School Certificate university entrance "Australian Tertiary Admission Rank" (ATAR) results in 2010 and

[3] Unfortunately, no such comparison was made between public and non-public schools at the primary level because there are no data available for non-public primary schools.

2008, (ii) the difference between schools' 2008 and 2010 median Year 10 School Certificate Exam result, and (iii) the difference between the average NAPLAN test scores for Year 9 in 2010 and the average NAPLAN test score for Year 7 in 2008. The authors also exploited the double bootstrap procedure for DEA as proposed by Simar and Wilson (2007), and found that schools with higher student retention rates, larger enrolments, single sex schools, and selective admissions schools tend to be more efficient than other schools. By contrast, schools from more remote areas, with a higher ratio of students from English as secondary language and Aboriginal backgrounds, high rate of students required special education were less efficient. They also found that the socio-economic background of students (proxied by ICSEA index) and experiences of the teachers had no statistically significant impact on the efficiency of schools.

Two recent papers have used national data to study the performance of schools across Australia. Miller and Voon (2012) used aggregate school level data and a regression/decomposition approach to document that the average test scores for independent schools were consistently the highest across the three sectors, while the scores for the government schools were the lowest. By contrast, Ryan (2013) used the Programme for International Student Assessment (PISA) data to examine academic achievements of 15 year-old secondary school students from 2003 to 2009. Using the PISA test scores in math and reading as outcomes in an SFA approach, he found that private schools had the highest efficiency scores, followed by Catholic schools and public schools. He also found that school level variables such as school autonomy measures, student–teacher ratios, information on admission practices and levels of school resources added little to the explanation of school performance.

In summary, there is quite a rich extant Australian literature on school performance. There is, however, scope to improve upon and add to this literature. In particular, with the exception of two studies which controlled for the quality of student intake either by including prior achievement as an input (Bradley et al., 2004) or using growth in scores as the output measure (Haug and Blackburn, 2013a), almost all Australian studies reviewed so far used the level of test scores as outcomes in their analyses and have thus been unable to effectively control for the quality of student intake, except via measures of socio-economic status. Furthermore, both of those studies are based on state-level data and focus exclusively on public schools. International (Elder and Jepsen, 2014; Hanushek and Taylor, 1990; Jepsen, 2003) and Australian (Bradley et al., 2004; Lamb et al., 2004; Marks, 2009; Nghiem et al., 2013) literature has shown that, due to problems associated with school selection, failure to account for the initial abilities of students may give rise to misleading conclusions about the relative performance of schools or the impact that school choices may have on student academic outcomes. This paper contributes to the literature by conducting a nationwide study of school efficiency which exploits the availability of data on the growth of NAPLAN test scores from the MySchool website. It thus effectively controls for the quality of student intake. Additionally, with the exception of two studies that use national data (Miller and Voon, 2012; Ryan, 2013) and the study by Lamb et al. (2004), all other Australian studies have focused on public schools. In this paper, we examine the performance of other school sectors, namely Catholic and independent schools, and compare the relative efficiency of various school types. By doing so, we make another contribution to the Australian literature on the impact of school choice on academic achievement. The Australian literature so far does not lead to a consensus view about which school types produce better academic outcomes for students (Vella, 1999; Marks et al., 2001; Marks, 2007, 2009; Ryan and Watson, 2010; Cardak and Vecci, 2013; Nghiem et al., 2013).[4]

3 Methodology

In this study we measure the efficiency of schools using the distance function concept pioneered by Shephard (1953) and Malmquist (1953). In particular, a distance function measures the distance from an actual observation of inputs and outputs to desired input-output combinations (i.e., combinations on the efficiency frontier). Following Coelli et al. (2005), we define the production technology T in which the relationship between a vector of inputs (x) and a vector of outputs (y) of a school is specified as:

$$T = \{(x, y) : x \text{ can produce } y\} \tag{1}$$

An output distance function, which measures the extent to which observed output can be expanded by a factor of $\theta \geq 1$ to reach frontier output keeping the amount of inputs unchanged[5] is defined as:

$$D_o(x, y) = Max\{(x, \theta y) \in T\} \tag{2}$$

The construction of the frontier can be undertaken parametrically (i.e., by undertaking a stochastic frontier analysis (SFA)) or non-parametrically (i.e., by conducting data envelopment analysis (DEA)). The parametric approach is similar to regression analysis with the exception that the error term has two components: a random error to capture noise and a non-negative component to represent efficiency.[6] By contrast, the non-parametric approach uses mathematical programming techniques to construct a piecewise frontier that envelops the data. The parametric approach is able to take into account random noise in the data but, conceptually, instituting it demands assumptions about economic behavior (e.g., that firms are profit-maximizing or, at least, cost-minimizing), along with distributional assumptions about the (in)efficiency component and its functional form.[7] An advantage of the non-parametric approach is that it does not require any such behavioral assumptions to hold; a drawback is that it

[4] See Nghiem et al. (2013), for example, for a review of this related literature.

[5] An input distance function is defined similarly.

[6] In some specifications such as cost function specifications, this component represents inefficiency.

[7] These assumptions have been demonstrated to exercise considerable influence over efficiency estimates and ranks. See, e.g. Street's (2003) demonstration of this in relation to hospital efficiency estimates.

does not take into account random noise. In practice, though, the effects of noise can be mitigated in non-parametric studies using a bootstrapping approach, and via the detection and removal of outliers.

In this study, we apply a non-parametric approach and use DEA with bootstrapping to study school efficiency. In essence, the output-oriented technical efficiency in DEA is measured by solving a linear programming problem below:

$$D_o\{x_i, y_i\} = max\{\theta_i|\lambda_i Y \geq \theta_i y_i;\, x_i \geq \lambda_i X,\, \theta_i \geq 1, \lambda_i \geq 0\} \tag{3}$$

where θ is a scalar representing technical efficiency ($\theta - 1$ represents the proportional increase in outputs that can be obtained while using the same level of inputs); λ is a vector of weights (also referred to as "peer weights") that represent the distances between a school and its peers; and Y and X represent the matrices of outputs and inputs, respectively, of all schools in the dataset. Empirically, the reciprocal of θ, which ranges from 0 to 1, is in fact used to measure the technical efficiency of the ith school. For example, a school with θ=1.25 means that the school could, if fully efficient, increase its output (e.g., test scores) by 25 per cent (1.25-1) using the same quantities of inputs and its reported technical efficiency score is hence $\frac{1}{1.25} = 0.8$, or 80 per cent.

In the extant literature many of the previous studies applied bootstraps to obtain the statistical properties of the resulting scores from DEA. Simar and Wilson (1998), though, have drawn attention to the fact that that most of the naive bootstraps that have historically been applied do not take into account the property that technical efficiency scores range from 0 to 1. Simar and Wilson (1998) overcome this problem by using a truncated distribution to redraw the sample.[8]

In this study, in order to take into account the effects of environmental factors (e.g., the socio-economic backgrounds of students, types of school) we, like, e.g. Chakraborty and Blackburn, 2013, apply the second-stage regression approach, which regresses the technical efficiency score estimated in the first stage against environmental variables in the second stage using the following equation:

$$y_i^* = z_i\beta + \varepsilon_i \tag{4}$$

where ε_i is a random error with mean zero and y_i^* is the latent variable that has the relationship with the observed technical efficiency as follow:

$$y_i = \begin{cases} y_i^* & if\, 1 > y_i^* > 0 \\ 1 & if\, y_i^* \geq 1 \\ 0 & if\, y_i^* \leq 0 \end{cases} \tag{5}$$

A problem with the second stage regression is that it violates an assumption of the classical linear regression model, namely that the dependent variable

[8] Note, however, that even the bootstrap procedure by Simar and Wilson (1998) did not actually take into account random noise or measurement error but it checks the sensitivity of results due to sample variability, hence, it is not necessary to conduct bootstrap for DEA studies that use complete population enumerations, such as census data (Coelli et al., 2005, pp.202-203).

are inter-independent. For example, when efficient schools are removed from the sample, the efficiency scores of the remaining schools will change as a new frontier will be estimated. Simar and Wilson (2007, pp.42–43) introduced a double bootstrap procedure to address this problem. To summarize, their double bootstrap procedure involves the following two main steps:

1. Construct the bias-corrected technical efficiency scores using original DEA estimates and re-sample the data using the original truncated regressions; and

2. Construct the bootstrapped confidence interval for the second stage regression using the bias-corrected technical efficiency score plus the non-negative noise drawn from a truncated distribution.

This estimation procedure is implemented in this paper using the FEAR package for the R statistical computing language (R Core Team, 2013) by Wilson (2008). The sampling procedure uses 1,000 repeats in both steps. In the test analysis with a small proportion, we tried to increase the number of repeats to 5,000 but found it made little difference to the confidence interval. Thus, for pragmatic reasons (i.e., to keep the calculation times manageable), we set the number of repetitions at 1,000 .

4 Data

4.1 Data source and variable selection

This study uses the National Assessment Program–Literacy and Numeracy (NAPLAN) data available from a website called "MySchool". NAPLAN, which was initiated in 2008, is a common national assessment program for all students at grade 3, 5, 7 and 9 in all Australian government and non-government schools.[9] Students are tested across five domains: Grammar and Punctuation, Numeracy, Reading, Spelling and Writing. In May of each year, all students in the same grade across Australia are assessed using the same tests. The tests are designed in such a way that the test results are all measured on common scales, ranging from 0 to 1000, rendering them comparable across grades. Specifically, the results of the tests are not only comparable across schools for the same classes in the same year, but also across school years (ACARA, 2013b). School-level aggregate results of these tests are made available to the public via the MySchool website in two main forms: the site reports score levels for all schools and also score growth. These are disaggregated by test year, grade, and domain. As at the time of this study, the MySchool website contains annual NAPLAN scores from 2008 to 2012 and score growth for three periods (2008-2010, 2009-2011, and 2010-2012). Besides information on NAPLAN test results, the MySchool

[9] Some students are, however, exempt from the tests. These include students from a non-English speaking background who have been in Australia for less than one year before the tests and students with substantial intellectual disabilities.

website also contains other school information including information about the school type (i.e., public, Catholic or other independent), the range of years taught at the school, the number of students enrolled, the number of teaching staff, the number of non-teaching staff, the recurrent income of the school, capital expenditure by the school, the proportion of students from non English-speaking backgrounds at the school and the proportion of children who identify as an Aborigine or Torres Strait Islander.

The main output measure we use is, as is discussed further below, the growth of NAPLAN test scores. For that reason, we restrict our empirical sample to schools with a sufficient number of students who took the tests in both 2009 and 2011.[10] We chose 2009-2011 period rather than the other alternatives (2008-2010 or 2010-2012) because it maximizes the number of observations in the dataset (i.e. not all schools have NAPLAN results for 2008 when the tests started) and the availability of input measures (the MySchool website contained no information on school finances for 2012 at the time of this study). We also exclude special schools that serve students with a disability from the analysis because students with a substantial intellectual disability are not required to take NAPLAN exams.[11] Finally, we exclude 23 schools with missing information on expenditure as well as 18 schools that are distance education (DE) providers.[12] Our final sample consists of precisely 6,778 schools.

We use the average growth in NAPLAN scores in Numeracy and Reading domains as the main school outputs of interest. This "unadjusted score growth" is calculated as the difference in test scores of the same students who took the NAPLAN tests at the same school in two different grades in 2009 and 2011.[13] We focus on the Numeracy and Reading domains of the NAPLAN as indicators of school output because MySchool does not report student gains in other NAPLAN domains (ACARA, 2013a). The use of score growth as a measure of school output is a significant improvement over most of the previous Australian literature which had been unable to control effectively for the quality of student intake at each school. Note that while this score growth measure provides an attractive way to examine school efficiency, it does not capture the characteristics of the learning process as well as the structure of the full NAPLAN results would. Furthermore, it is also important to bear in mind that all NAPLAN results are to be regarded as indicators of the latent variable of interest. It is possible, too, that the rate at which current learning builds on past performance varies along the NAPLAN score distribution (Hanushek et al., 2007; Grosskopf

[10]According to ACARA, score growth is shown only for schools with five or more matched students.

[11]In addition, ACARA excludes these schools when calculating average for schools with similar backgrounds or same starting scores.

[12]Our rationale for excluding DE providers is that these schools are liable to have fundamentally different production processes than the majority of schools in our sample.

[13]Note that the study by Haug and Blackburn (2013b) which also uses NAPLAN data defines score growth differently. In particular, they calculate score growth as the difference between the average scores in two years for the same cohort (not the same students). As they note, their score growth measure does not take into account the possibility that students might change schools over time.

et al., 2009). For instance, it may perhaps be easier to improve the NAPLAN scores of students who start from a lower score base than to improve from a high base (or vice-versa). We address the latter concern by comparing the changes in NAPLAN results of students with the same starting level. In particular, we follow some of the US literature (Reback, 2008; Gronberg et al., 2012; Grosskopf et al., 2009) to measure a school's output as its deviations from the expected score of the schools with the same previous test scores. We name this output the "adjusted score growth" to distinguish it from the "unadjusted score growth" concept that was introduced previously. For the purpose of comparing our results with those of other studies (Miller and Voon, 2012) that use the same data set but different measures of output, we also use the average of scores in Numeracy and Reading domains as the output in the DEA model.

The input measures that are used in this study are the teacher/student ratio, non-teaching staff/student ratio, recurrent income per student and capital expenditure per student.[14] The first three inputs and other environmental variables introduced below are measured each year, but for capital expenditure per student we take the average of capital expenditure over the entire 2008 - 2011 period to smooth out large expenditures in any given year. To measure the impact of environmental variables on school efficiency we include the Index of Community Socio-Educational Advantage (ICSEA) which represents levels of educational advantage in the second stage regression. This index, which takes values from approximately 500 to approximately 1300, is constructed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) taking information relating to education and occupation of parent/guardian, geographical remoteness and the proportion of Aboriginal and Torres Strait Islander (ATSI) enrollments of the school into account (ACARA, 2013c). By construction, schools with higher ICSEA scores are schools with more students from educationally advantaged backgrounds. We also include in the regression dummy variables for girls-only schools and boys-only schools to represent the characteristics of single sex schools and enable their comparison with co-educational schools. We also include the ratio of non-English speaking background students as an indicator of linguistic and cultural background that is not captured by the ICSEA. Other environmental variables selected for the second stage regressions include: school type (i.e., public, Catholic and other independent schools, taking public schools as the base), school year level (mixed level or secondary schools, primary schools are selected as the base), state (dummies for states and territories with NSW/ACT is selected as the base), the number of students (and its square). Finally, while the proportion of Aboriginal and Torres Strait Islander (ATSI) enrollments and school geographical remoteness are included in the calculation of ICSEA. Given their small weights in the ICSEA calculation, we also include them as environmental variables to test for any additional impact these variables may have on school efficiency. [15]

[14]Note that the use of the recurrent income measure involves, in the production context, the inherent assumption that recurrent expenditure equals recurrent income.

[15]We tried to collect information about teachers' characteristics but unfortunately this type of information is not available from the MySchool website. Missing information on teachers'

4.2 Descriptive statistics

The descriptive statistics of the inputs, outputs and environmental variables (Table 1) show that the average NAPLAN score of students in Australia in 2011 is 440, and the average unadjusted (adjusted) growth of test scores between 2009 and 2011 is 0.36 (0.34) points. One notable characteristic from the descriptive statistics is that the minima for NAPLAN test growth, both adjusted and unadjusted, are negative. Negative values for output growth violates the basic assumption in DEA that the output measure cannot be negative. To deal with this, we thus convert the growth of outputs to be positive integers by subtracting the minimal value and adding an arbitrarily positive value of 10 (i.e., the minimum value of the transformed output is now 10). The data also show that, on average, schools in Australia have 68 teachers and 25 non-teaching staff per 1,000 students. The average capital expenditure and income of schools is A$11,000 and A$12,000 per student per year, respectively. Note that the data on capital expenditure fluctuates considerably over time due to unexpected large spending in some years (i.e., the purchase of new equipment or the construction of new buildings). Thus, we take the average of capital spending over the 2008-2011 period to smooth out the temporal variation that arises due to the temporal lumpiness of investments in school capital. The descriptive statistics show that about 69 per cent of schools in Australia are public, with Catholic schools making up 20 per cent and the remaining 11 per cent being made up of other independent schools.

Primary schools account for 72 per cent of the sample, while secondary schools account for 15 per cent of schools and the remaining schools are mixed-level schools that combine primary- and secondary-level school years. Among the eight states and territories of Australia, NSW and the ACT (combined) account for the largest proportion of Australian schools, accounting for 36 per cent of schools in our sample, followed by Victoria (25 per cent) and Queensland (16 per cent).[16] The Northern Territory has the lowest share of Australian schools (1.6 per cent) and single-sex schools represent a similar proportion of all Australian schools. On average, 19 per cent of students in Australian schools come from a Non-English Speaking Background (NESB) and 6.7 per cent of students identify as Aborigines or Torres Strait Islanders (ATSI).[17]

characteristics may not have a large effect on our findings: previous studies such as Leigh (2010), Buddin and Zamarro (2009), Loman et al. (2012), and Haug and Blackburn (2013a) have found that teacher characteristics including educational background and salary explain only a small fraction of the variations in the effects of teachers on improving students' test scores. It is also worth noting that the estimated parameters on ICSEA are insensitive to the inclusion of ATSI ratio and remoteness in the second stage regression.

[16]Due to the small number of schools in the Australian Capital Territory (ACT) that had NAPLAN test data in both 2009 and 2011, we combined the ACT with New South Wales (NSW), one of its geographical neighbors.

[17]In the regressions analyzes we divided the NESB and ATSI rates by 100 and the ICSEA index by 1000 to make their scale in the range with other covariates. This has no effect on the relationship as regressions are independent of measurement unit but it renders the parameter estimates visible at three decimal points.

Table 1: Descriptive statistics

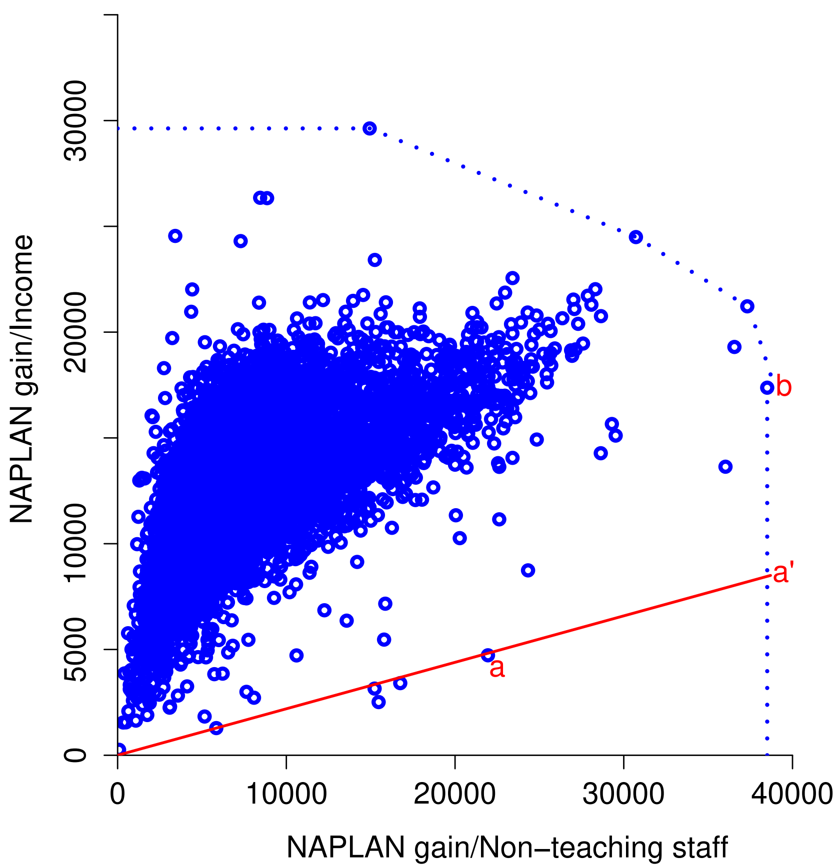
	Means	Std	Min	Max
Outputs				
NAPLAN score	440.614	59.137	84.000	745.000
Unadjusted growth	0.359	23.187	-185.000	157.500
Adjusted growth	0.340	15.830	-127.250	130.500
Inputs				
Teacher/student (persons)	0.068	0.017	0.018	0.344
Non-teacher/student (persons)	0.025	0.015	0.004	0.286
Expenditure/student ($mil.)	0.011	0.004	0.004	0.099
Income/student ($mil.)	0.012	0.004	0.004	0.090
Environmental variables				
Number of students (persons)	430.559	339.140	23.000	3271.000
ICSEA index	1.010	0.094	0.511	1.239
NESB rate	0.193	0.242	0.000	1.000
ATSI rate	0.068	0.134	0.000	1.000
Boy schools (1=yes)	0.017	0.129	0.000	1.000
Girl schools (1=yes)	0.021	0.145	0.000	1.000
Public schools (1=yes)	0.685	0.465	0.000	1.000
Catholic schools (1=yes)	0.206	0.404	0.000	1.000
Private schools (1=yes)	0.109	0.312	0.000	1.000
Primary schools (1=yes)	0.726	0.446	0.000	1.000
Mixed schools (1=yes)	0.123	0.329	0.000	1.000
Secondary schools (1=yes)	0.151	0.358	0.000	1.000
Metropolitan schools (1=yes)	0.613	0.487	0.000	1.000
Provincial schools (1=yes)	0.347	0.476	0.000	1.000
Remote schools (1=yes)	0.040	0.197	0.000	1.000
NSW & ACT	0.359	0.480	0.000	1.000
Victoria	0.254	0.435	0.000	1.000
Queensland	0.160	0.366	0.000	1.000
Western Australia	0.105	0.306	0.000	1.000
South Australia	0.075	0.264	0.000	1.000
Tasmania	0.032	0.176	0.000	1.000
Northern Territory	0.015	0.124	0.000	1.000

An empirical frontier, drawn using the ratio of output on two of the four inputs (i.e., recurrent expenditure and non-teaching staff), shows that the frontier includes only a very small number of schools. Additionally, only a small number of schools with high output-to-input ratios would constitute relevant peers for most of the inefficient schools. Figure 1 also suggests that many schools will have slacks in inputs/outputs.[18] For example, school a in Figure 1 could be

[18]Slacks refers to congestion in inputs such that firms would be able to produce more output with less input. In the context of schools, factors like workplace regulations or the unionization of labor may prevent schools from using more efficient combinations of inputs than could otherwise be chosen. For more detailed discussion about slacks, see, for example,

technically efficient if it achieved the input/output combination at a', but it could improve its NAPLAN growth even further by moving towards the position of school b, for instance. Thus, for this inefficient school, the distance aa' is the output gained by improving technical efficiency whilst $a'b$ is the output growth that could be achieved by solving input congestion. The other empirical frontiers, sketched using other combinations of four inputs depict a very similar story but we present only one frontier here, for brevity.

Figure 1: An example of an empirical frontier



Note: Adjusted NAPLAN growth (gain) is used as the outcome in the plot.

Coelli et al. (2005).

5 Results and discussion

5.1 Main model

Table 2 presents a summary of the technical efficiency scores obtained from 48 DEA models disaggregated by school levels, school sectors and choices of outputs.[19] The results show that schools of the same type (i.e., public, Catholic and independent schools) and levels (i.e., primary, mixed and secondary schools) are likely to operate under the same technology. The average technical efficiency scores under these separate frontiers are always higher than those obtained from an aggregated frontier because the disaggregated frontier forms a tighter "envelope" around the data when schools are delineated by type. On these separate frontiers, independent and Catholic schools are also always more efficient than their public counterparts. The choice of output, though, seems to change the average technical efficiency scores of schools. Compared with the average technical efficiency scores of 75 per cent for primary schools and 89 per cent for secondary schools reported in the study by Chakraborty and Blackburn (2013), our estimates are considerably lower. Besides the differences between the dataset used in that study and our, and the selection of different inputs, the choice of output appears to be the main reason behind these differences in estimated efficiency scores. Specifically, we use only one NAPLAN dimension at a time while they use several outputs (e.g., test scores for different subjects). When the dimension of output comparison increases the number of schools on the frontier will increase and the distance from the frontier to inefficient schools will be reduced, hence generating higher efficiency scores. For example, schools that perform well on one subject (e.g., Numeracy) and poorly on other subjects will still be considered as efficient if multiple subjects are selected as outputs, while this will not necessarily occur when a single output is selected for analysis. We believe that achievement across all subjects using total or average scores is likely to be a better measure of school performance, especially at primary or secondary levels. On the other hand, using the output measures we chose for this study does enable us to consider a national sample and to make a number of comparisons that have not been possible in the extant Australian literature.

To explore the influence on efficiency scores of using an adjusted or unadjusted growth measure, we conducted pairwise correlations between efficiency scores using both measures. The correlation analyses showed that technical efficiency scores estimated by using the unadjusted and adjusted growth of NAPLAN test scores are highly correlated. For instance, when all schools are included in the analysis, the correlation coefficient of technical scores when adjusted and non-adjusted growth are used as outputs is 0.67. Nevertheless, we believe that the adjusted growth measure is conceptually the most suitable output measure because it takes into account the possibility that schools with higher initial scores face more difficulty in obtaining the same score growth as

[19] Because all inputs and outputs in this study are at ratio formats (e.g., teacher/student ratio and average growth of NAPLAN test score), the role of scale is neutralized, and hence, we focus on discussing technical efficiency in this paper.

Table 2: Technical efficiency by school types, school levels and output choices

		Output1		Output2		Output3	
	N	Mean	Std.	Mean	Std.	Mean	Std.
All schools	**6778**	**0.66**	**0.09**	**0.62**	**0.11**	**0.59**	**0.10**
Public	4642	0.68	0.12	0.66	0.11	0.63	0.10
Catholic	1394	0.83	0.09	0.70	0.12	0.70	0.13
Independent	742	0.84	0.07	0.74	0.08	0.71	0.08
Primary schools	**4919**	**0.75**	**0.08**	**0.65**	**0.12**	**0.61**	**0.11**
Public	3678	0.74	0.08	0.68	0.12	0.64	0.11
Catholic	1112	0.83	0.07	0.72	0.12	0.72	0.13
Independent	129	0.89	0.09	0.81	0.13	0.82	0.11
Mixed schools	**834**	**0.75**	**0.08**	**0.81**	**0.06**	**0.85**	**0.05**
Public	612	0.75	0.08	0.81	0.06	0.86	0.05
Catholic	183	0.95	0.04	0.89	0.05	0.90	0.04
Independent	39	0.94	0.07	0.90	0.07	0.92	0.06
Secondary schools	**1025**	**0.81**	**0.11**	**0.72**	**0.09**	**0.82**	**0.10**
Public	352	0.79	0.13	0.73	0.13	0.78	0.13
Catholic	99	0.88	0.10	0.87	0.07	0.92	0.08
Independent	574	0.85	0.06	0.76	0.07	0.87	0.06

Notes: Results from the first stage of the double bootstrapping DEA models. Inputs in all models are: teacher/student, non-teaching staff/student, capital expenditure/student, and income/student. Output1=level NAPLAN test scores, Output2=Unadjusted growth, Output3=Adjusted growth.

schools with lower initial schools. In the remainder of this paper we thus focus on examining the operational efficiency of Australian schools using the adjusted NAPLAN score growth as our output measure.

As mentioned previously, we apply the bootstrap procedure developed by Simar and Wilson (2007) to explore the determinants of schools' technical efficiency scores. When all schools are included to construct the frontier, the results (see Table 3) show that schools with more favorable socio-educational conditions (i.e., higher ICSEA) are more efficient. The magnitude of the ICSEA index is also the largest of the parameters, suggesting that it is the dominant determinant of school efficiency. For example, a school with the most advantageous socio-educational background (i.e., ICSEA index=1.3) is more efficient than the school with least the least advantageous (i.e., ICSEA=0.5) by $.492 \times (1.3 - .5) \times 100 = 39.4$ percentage points. Other indicators such as the school's remoteness, the proportions of students identifying as ATSI or from a NESB suggest that schools with disadvantageous backgrounds are more efficient, other things equal. For example, one percentage point increase in the ATSI rate is associated with a $.01 \times .086 \times 100 = .09$ percentage point increase in technical efficiency. That higher ATSI rates are associated with greater efficiency seems

Table 3: The determinants of school efficiency - results from regressions of all schools

	Adj. growth	Unad. growth	Level scores
Catholic schools	$-.019$ (.001)***	$-.006$ (.001)***	$-.009$ (.002)***
Private schools	$-.062$ (.003)***	$-.004$ (.002)*	$-.054$ (.003)***
ICEAS index	.492 (.011)***	.479 (.009)***	.304 (.013)***
NESB rate	.023 (.003)***	$-.020$ (.003)***	.017 (.004)***
ATSI rate	.086 (.008)***	.048 (.006)***	.113 (.009)***
Boys only	$-.033$ (.005)***	.007 (.004)*	$-.040$ (.006)***
Girls only	$-.033$ (.004)***	.013 (.004)***	$-.029$ (.005)***
Mixed schools	$-.067$ (.002)***	.128 (.002)***	$-.083$ (.003)***
Secondary schools	$-.048$ (.002)***	.043 (.002)***	$-.051$ (.003)***
Provincial schools	.002 (.002)	.004 (.001)**	$-.006$ (.002)**
Remote schools	.023 (.004)***	$-.001$ (.003)	.018 (.004)***
Log of students	.073 (.008)***	$-.036$ (.007)***	.096 (.010)***
Log of students squared	$-.004$ (.001)***	.004 (.001)***	$-.006$ (.001)***
Victoria	$-.044$ (.001)***	.008 (.001)***	$-.024$ (.002)***
Queensland	$-.092$ (.002)***	.018 (.001)***	$-.081$ (.002)***
Western Austrlia	$-.104$ (.002)***	.015 (.002)***	$-.105$ (.003)***
South Australia	$-.075$ (.002)***	.021 (.002)***	$-.068$ (.003)***
Tasmamia	$-.096$ (.003)***	.003 (.003)	$-.077$ (.004)***
Northern Teritory	$-.073$ (.005)***	$-.027$ (.004)***	$-.070$ (.006)***
Constant	$-.408$ (.025)***	.093 (.021)***	$-.245$ (.031)***

*** p<0.01,** p<0.05, * p< 0.1. Standard errors are in parentheses.

to be in contrast with the findings of Chakraborty and Blackburn (2013) but their results apply to secondary schooling, which our results also confirm when this group is investigated separately (see Table 4). Table 4 also shows that both Catholic and independent schools are less efficient than public schools when growth measures of output are used. When levels of NAPLAN test scores are used as an output measures, though, this result is reversed. This illustrates the importance that the specification of the output measure can make to the result. Again, we believe that the growth of NAPLAN test score is a better output measure because differencing the scores obtained by the same student in two consecutive tests removes the time-invariant unobserved characteristics of students that may otherwise distort comparisons. By location, schools in all other States and Territories are less efficient than those in NSW and ACT. We also find that single-sex schools (i.e., boys or girls schools) are less efficient than unisex schools, and mixed level schools and secondary schools are less efficient than primary schools.

Some of the results reported in Table 4 would seem to be counter-intuitive: schools in more remote locations and schools that have higher ratios of ATSI students –a group that is known, on average, to be disadvantaged– are more effi-

Table 4: The determinants of school efficiency - results from separate regressions by school types

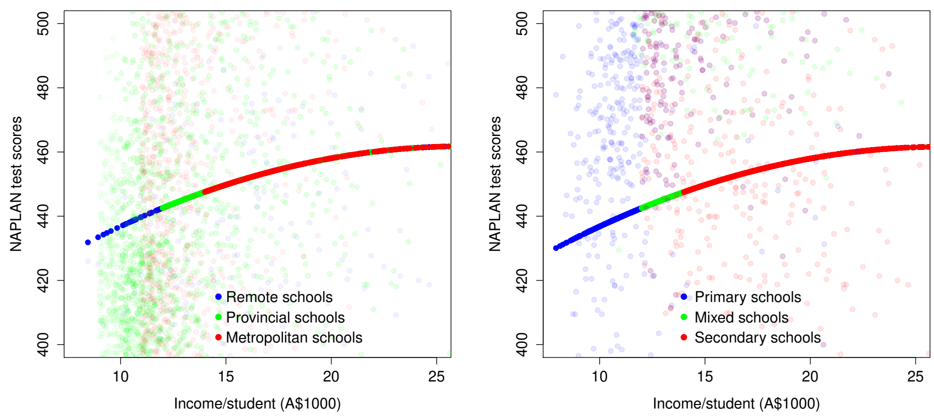
	Primary	Mixed	Secondary
Catholic schools	$-.009$ (.002)***	$-.001$ (.003)	.073 (.006)***
Private schools	$-.036$ (.006)***	.024 (.006)***	.079 (.005)***
ICSEA index	.588 (.013)***	.351 (.023)***	$-.208$ (.038)***
NESB rate	.000 (.004)	.022 (.005)***	$-.002$ (.010)
ATSI rate	.073 (.011)***	.036 (.023)	$-.311$ (.019)***
Boys only	$-.231$ (.046)***	.012 (.005)**	$-.053$ (.007)***
Girls only		$-.015$ (.004)***	$-.079$ (.008)***
Provincial schools	.007 (.002)***	.013 (.003)***	$-.005$ (.005)
Remote schools	.021 (.005)***	$-.011$ (.013)	$-.020$ (.007)**
Log of students	$-.003$ (.013)	.086 (.039)*	.178 (.023)***
Log of students squared	.004 (.001)***	$-.006$ (.003)*	$-.014$ (.002)***
Victoria	$-.032$ (.002)***	.005 (.003)*	.008 (.005)
Queensland	$-.087$ (.002)***	.046 (.030)	.050 (.005)***
Western Australia	$-.099$ (.003)***		.045 (.005)***
South Australia	$-.080$ (.003)***	.009 (.030)	.028 (.006)***
Tasmamia	$-.101$ (.005)***	.013 (.006)*	.034 (.008)***
Northern Teritory	$-.109$ (.007)***	.011 (.010)	.065 (.011)***
Constant	$-.308$ (.036)***	.095 (.125)	.133 (.080)*

Note: *** p<0.01,** p<0.05, * p< 0.1. Standard errors are in parentheses. Some cells are blank due to missing data.

cient. Reasons for these apparently counter-intuitive results may be diminishing marginal returns in the production function as well as our choice of output measure, i.e. the gain in NAPLAN test score between 2009 and 2011. To explore this question we present an empirical in Figure 2, where we have fitted a second order polynomial model to illustrate a production function using one input (income/student) and one output (test scores). Figure 2 shows that marginal return on test scores is decreasing in the quantity of the input variable. In particular, Figure 2(a) suggests that the reason remote schools are, on average, more efficient than metropolitan schools is because they operate on a section of the production function that has a reasonably steep gradient. Many metropolitan schools, by contrast, appear to be operating close to the flat-of-the-curve. As a result, the marginal test score growth of remote schools may be higher, on average, than that of metropolitan schools. Figure 2(b) paints a similar picture to that of previous studies such as Daraganova et al. (2013) and Nghiem et al. (2013): NAPLAN test scores generally increase with grade–test scores of secondary schools are higher than those of primary schools and many secondary schools will operate near the flat-of-the curve of the production function.

Figure 2: Empirical production functions

(a) By geographical locations (b) By school types



As expected, when each type of schools is examined separately (see Table 4), the signs of environmental variables change considerably depending upon whether primary or secondary schools are examined. In particular, schools in remote areas and schools with high proportions of students from ATSI background are less efficient at the secondary level although they are deemed to be more efficient at the primary level. Also, at the primary level Catholic and independent schools are less efficient than public schools but the story is reversed both for mixed and secondary-only schools. Similarly, schools in all other states and territories are less efficient than those in NSW and ACT at the primary level but they become more efficient than NSW and the ACT in mixed-level schools and secondary schools. The parameters of the number of students and its quadratic term are positive and negative, respectively, suggesting scale economies are present at mixed and secondary level schools. The statistically insignificant results on these coefficients for primary schools suggest that scale economies may not be important to NAPLAN score growth in primary schools; there is also no evidence of scale diseconomies for primary schools. Finally, with the noteworthy exception of boys-only schools of mixed levels, all other single-sex schools are less efficient than coeducational schools at all levels. Our finding on the effect of gender composition of schools on efficiency for secondary schools is in contrast with those found by Haug and Blackburn (2013a). The latter study found that single-sex schools are more efficient than unisex schools and that the socio-economic background of parents had no significant effect on school efficiency. Again, our choice of different outputs to theirs–recall that they used the median tests scores of Higher School Certificate for Year 12, School Certificate exam for Year 10, and average NAPLAN test scores for Year 7 and

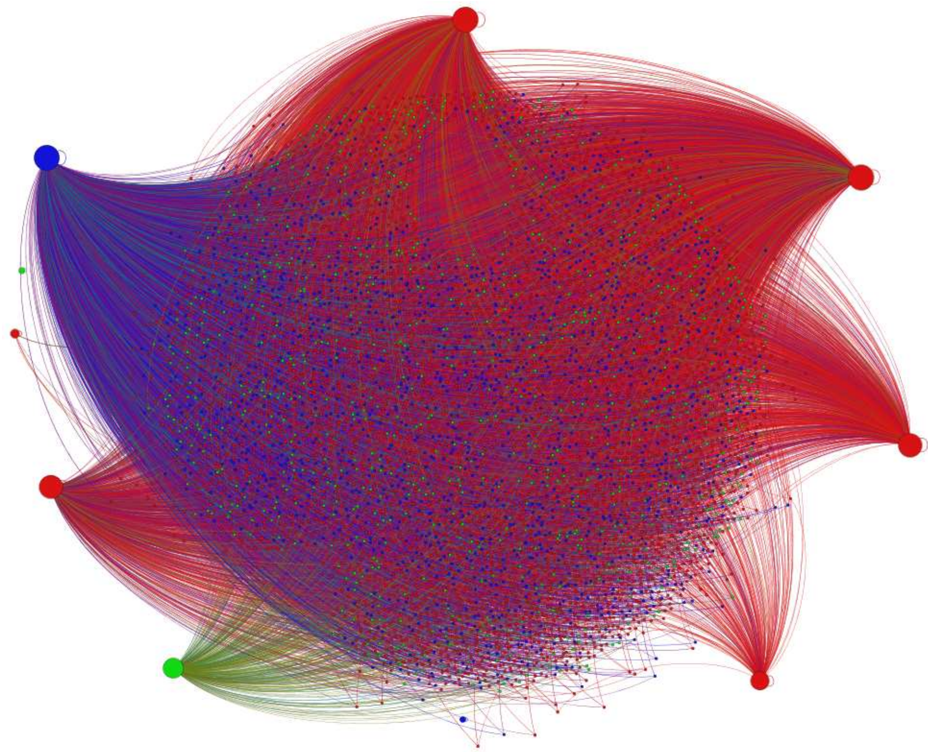
Year 9–probably explain this difference.[20] Our results for secondary schools are, however, in-line with those of Lamb et al. (2004) who found that, at the secondary level, Catholic and independent schools are more efficient than their public counterparts. Similarly, the finding that primary level public schools perform better than Catholic and independent schools is also in line with the recent findings of Nghiem et al. (2013).

DEA results can also provide useful information about peers and peer weights (for inefficient schools), and peer counts (for efficient schools). As was mentioned previously, peers refer to efficient schools that hence have an input-output structure that inefficient schools can learn from to improve their operational efficiency. When more than one peer is available to an inefficient school, the relative importance of each peer is based on its peer weights (i.e., vector λ in Equation 3). For efficient schools, the number of peer counts refers to the number of other schools referred that may be regarded as production "role models"; schools with bigger peer counts are hence more influential. We summarize the peer information obtained from DEA in Figure 3 using Gephi version 0.82 (Bastian et al., 2009). We present the size of each school as nodes in Figure 3 using the logarithm of their peer counts for ease of viewing, because only schools with very large number of peer counts—the highly influential schools–are visible. The remaining schools, including efficient schools that have relatively few peer counts, are hardly visible. Types of schools in Figure 3 are represented by different colors: red denotes public schools, blue denotes Catholic schools and green denotes other independent schools. The network graph in Figure 3 confirms earlier observations from the empirical frontier in Figure 1 that only a few schools are highly influential in the sense that they are both fully efficient and have large peer counts. The summary of connectedness of the network shows a high level of connectivity with three lines (also known as edges) per school, on average. Also, no school in the map stands alone (i.e., is without an edge), suggesting that no school has an input/output structure so unique that they cannot be compared with other schools. This is not particularly surprising given the relatively high degree of regulation of this sector. It is interesting that most of influential schools are public sector, while Catholic and independent schools have one influential school each. The over-representation of public schools in the list of influential schools is reasonable because public schools comprise more than 70 per cent of all schools. The map also shows that inefficient schools are more commonly of the same type: the rays from influential schools in Figure 3 more often than not have the same color as the influential schools. This result is also to be expected because schools of the same type may also be expected to share similar input/output structures and because DEA constructs targets for inefficient schools based only on peers that have similar such structures. One exception is the influential Catholic school represented by the big blue node on the left of Figure 3. It also plays the role of a peer to inefficient schools from the public and private sectors. It is also remarkable that most of the influential

[20]The authors of that paper also had a considerably smaller sample size (i.e., 380 public schools in NSW) to work with.

schools on the map are from NSW/ACT and Victoria.

Figure 3: Peers network



The summary of projected inputs and outputs presented in Table 5. This summary shows that, on average, schools have the potential to gain 237 NAPLAN points between 2009 and 2011. Note that this improvement is computed on the output that was rescaled to render it positive for the purposes of applying DEA. Converting the gain to the original scale, the projected output is 100 points (i.e., adding the minimum growth in Table 1 and subtracting 10),[21] which is still a remarkably large improvement compared with the original average of 0.33 points. One may wonder why the projected improvement is so substantial. Again, the reason is because we have converted the selected output (i.e., the adjusted growth of NAPLAN scores) to be positive by shifting the output distribution to the right so that the minimum gain is 10. Thus, the average of transformed score is approximately 140 and hence the projected output is 237 (i.e., 140×1.695). Table 5 also shows that the average projected inputs are slightly smaller than the data reported in Table 1. In particular, as

[21]The projected output is 237, and the minimum score in the sample is -127 (Table 1), hence the transformed projection is computed as: 237-127-10=100.

Table 5: Summary of targets and slacks

	Mean	Std.	Min	Max
Test score growth (NAPLAN scores)	236.50	25.57	124.42	267.75
Teachers/student	0.056	0.006	0.018	0.072
Non-teaching staff/student	0.023	0.010	0.004	0.037
Capital expenditure/student ratio (A$million)	0.009	0.001	0.004	0.015
Income/student (A$million)	0.010	0.001	0.004	0.015

the teacher ratio decreases from 68 to 56 teachers per 1,000 students, the ratio of non-teaching staff decreases from 25 to 23 persons per 1,000 students, and average capital expenditure per student decreases from A$11,000 to A$9,000; the relative reduction of income per student is from A$12,000 to A$10,000. Although we choose an output-oriented approach to measure school efficiency, the reduction of inputs in targets suggests that some schools have input slacks or congestion. One possible reason for input congestion may include government regulations pertaining to maximum class sizes and structural aspects of the Australian education sector. Leigh and Ryan (2006), for instance, point to labor regulations, the prevalence of unionization, and the rejection of merit-based pay by the union movement as structural factors that may make it costly or difficult for Australian schools to discipline or remove teachers and non-teaching staff to achieve operational efficiency. Regarding the last input, too much spending on the "right" items or spending on "wrong" items may also cause adverse effects on the NAPLAN score gain. Overall, due to input congestion, on average schools employ 12 more teachers and two more non-teaching staff per 1,000 students than is estimated to be optimal, resulting in A$2,000 more expenditure per student above the optimum. Given that there were approximately three million students in the schools included in this study, this result suggests sectoral reductions of 36,000 fewer teachers, 6,000 fewer non-teaching staff, and A$12 billion less per annum in capital expenditure and recurrent spending could, in theory, be achieved. These estimates represent the order of magnitude of estimated inefficiencies in the Australian schooling sector and seem unlikely to be realized without regulatory and structural reform in the sector.

5.2 Sensitivity analyses

To test the robustness of our results, we tried three other specifications of the model (see Table 6). First, we took further measures to test for the possibility that schools with lower starting scores are more likely to have higher NAPLAN growth due to a ceiling effect. For instance, a school with a test scores of 500 in 2009 potentially could gain up to 500 points (recall that NAPLAN test scores range from 0 to 1000), while the maximum gain possible for schools that start with a test score of 600 in 2009 is only 400 points. Hence, the chance

Table 6: Sensitivity analysis

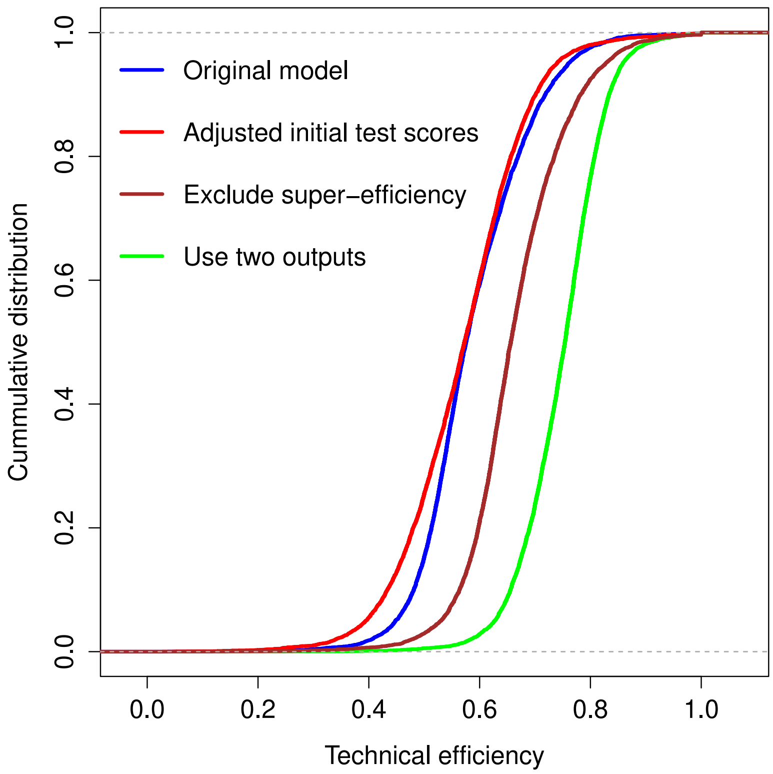
	Adjusted for initial scores	Use two outputs	Drop super-efficient schools
Catholic schools	−.015 (.001)***	−.011 (.001)***	−.011 (.002)***
Private schools	−.050 (.003)***	−.014 (.002)***	−.025 (.003)***
ICEAS index	.773 (.010)***	.553 (.008)***	.499 (.012)***
NESB rate	.012 (.003)***	−.005 (.002)*	.022 (.003)***
ATSI rate	.155 (.007)***	.052 (.006)***	.063 (.008)***
Boys only	−.016 (.004)***	.001 (.003)	−.029 (.006)***
Girls only	−.023 (.004)***	−.004 (.003)	−.032 (.005)***
Mixed schools	.050 (.002)***	.083 (.001)***	−.062 (.002)***
Secondary schools	.001 (.002)	.033 (.002)***	−.040 (.002)***
Provincial schools	.006 (.001)***	.007 (.001)***	.005 (.002)**
Remote schools	.011 (.004)**	.011 (.003)***	.017 (.004)***
Log of students	.008 (.008)	.010 (.006)*	.050 (.009)***
Log of students squared	.001 (.000)*	.001 (.000)*	−.003 (.001)***
Victoria	−.011 (.002)***	.001 (.001)	−.008 (.001)***
Queensland	−.034 (.002)***	.005 (.002)***	−.018 (.002)***
Western Austrlia	−.045 (.002)***	−.006 (.001)***	−.019 (.002)***
South Australia	−.027 (.002)***	.001 (.002)	−.038 (.003)***
Tasmamia	−.054 (.003)***	−.014 (.002)***	−.033 (.004)***
Northern Teritory	−.064 (.005)***	−.044 (.004)***	−.023 (.006)***
Constant	−.521 (.024)***	−.052 (.019)**	−.234 (.027)***

*** p<0.01,** p<0.05, * p< 0.1. Standard errors are in parentheses.

of achieving higher growth may be better, *ceteris paribus*, among schools with lower test scores in 2009. To explore this possible effect we conduct a sensitivity test by creating a ratio of the average initial NAPLAN scores of the schools with the maximum scores in the sample as a weight for NAPLAN gain. Thus, most schools will have their output weighted down with the exceptions of those that obtained maximal scores in 2009. The effect of applying this specification to the average technical efficiency scores is not clear, *a priori*: efficiency scores will be higher if this ratio approach when discrepancies among schools decreases; but most schools have their output weighted down so it is unclear whether these two effects combined will raise or lower the efficiency scores of schools that are not on the frontier.

Second, we used the NAPLAN scores themselves as a second output together with the NAPLAN growth measure. We expect higher efficiency scores to be estimated with this model because the frontier is constructed from combinations of schools with high NAPLAN gains or high NAPLAN scores, or both. Finally, we remove super-efficient schools, which are also the most influential schools in this study, from the frontier.[22] We also expect this to result in a frontier that forms a tighter envelope around the data and hence that the average efficiency scores of inefficient schools will increase in this specification.

Figure 4: Distributions of technical efficiency scores from different specifications



The results of these sensitivity analyses are as expected: the technical efficiency scores increase when both the level and growth of test scores are selected as outputs and when super-efficient schools are excluded. However, when the gains in NAPLAN scores are weighted by the ratio of actual and maximum scores, the average technical efficiency decreases slightly. The cumulative distributions of technical efficiency scores in Figure 4 show that this transformation

[22]The super-efficiency score is calculated by constructing the frontier without one school on the frontier then projected their original output to the new frontier and the super efficiency score is the ratio of projected output and original output. A school is considered super-efficient if its super-efficiency score is less than one, meaning that they remain on the frontier even when their output are reduced (Coelli et al., 2005). In the literature, sensitivity is done by removing super-efficient firm one-by-one but we found this has little impact with our data. Also, because we have only a few super-efficient schools, we decide to remove all these schools in the sensitivity analysis.

in output increases the proportion of schools with efficiency scores lower than 50 per cent, whilst decreasing the proportion of schools with scores above 50 per cent. It also shows clear dominance of technical efficiency scores when super-efficient schools are excluded and when both level and growth of test scores are used as outputs over the distribution of efficiency scores in the remaining three scenarios of single output choices. The correlation coefficients also show strong correlations between the technical efficiency scores from the original specification and the first two sensitivity tests (i.e., adjusting for initial scores and excluding super-efficient schools). The final test (i.e., using two outputs) exhibits a weak correlation with the other models, which is to be expected because none of the single output models constructs a frontier from the linear combinations of level and growth of test scores.

Results from the second stage regressions produce evidence of little sensitivity among the three alternatives outlined above (see Table 6). In particular, schools with high ICSEA and high rates of NESB students are still more efficient. Also, Catholic and independent schools remain less efficient than public schools in all three sensitivity tests. The relative performance of schools from different States and Territories relative to NSW and the ACT also unchanged except when two outputs (both growth and level scores) are used. Similarly, single-sex schools are less efficient than other schools except when two outputs are used to benchmark schools.

6 Conclusions

This paper has examined the efficiency of Australian schools using data from the MySchool website. Unlike most previous Australian studies, we used the average growth of NAPLAN test scores of the same students as our primary measure of school output. We applied the double bootstrapped procedure of Simar and Wilson (2007) to examine the impact of environmental factors on the operational efficiency of schools. Our results show that the average technical efficiency of Australian schools is 59 per cent and congestion exists for all of the available inputs. If all schools are able to learn from their peers on the frontier and input congestion is controlled for, the average growth of test scores can be increased by 100 NAPLAN points whilst capital expenditure and income can be reduced by A$2,000 per student per year. In addition, on average, schools with 1,000 students can employ 12 fewer teachers and 2 fewer non-teaching staff if input congestion can be eliminated. For structural reasons, though, such as class size limits and labor market rigidities, some of these inefficiencies may be difficult to overcome. The socio-educational factors that are positively associated with school efficiency are the ICSEA and the proportion of students from an NESB background. On average, Catholic and independent schools are less efficient than public schools at the primary level. This relative performance by sector may hold for primary schools only since at higher school levels, Catholic and independent schools are more efficient than public ones. The results also suggest that the frontier was constructed of a small number of super-efficient schools.

These schools represent examples from which lessons for inefficient schools may be drawn.

References

ACARA (2013a): "My School Student Gain," Report, The Australian Curriculum, Assessment and Reporting Authority.

——— (2013b): "Reliability and validity of NAPLAN," .

——— (2013c): "Report on the generation of the 2012 Index of Community Socio-educational Advantage (ICSEA)," .

Agasisti, T. (2011): "The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data," *Education Economics*, 21, 520–544.

Alexander, W. R., A. Haug, and M. Jaforullah (2010): "A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools," *Journal of Productivity Analysis*, 34, 99–110.

Bastian, M., S. Heymann, and M. Jacomy (2009): "Gephi: an open source software for exploring and manipulating networks," in *International AAAI Conference on Weblogs and Social Media*.

Battese, G. E. and T. J. Coelli (1995): "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics*, 20, 325–332.

Blackburn, V., S. Brennan, and J. Ruggiero (2013): "Measuring efficiency in Australian Schools: A preliminary analysis," *Socio-Economic Planning Sciences*.

Bradley, S., M. Draca, and C. Green (2004): "School Performance in Australia: Is There a Role for Quasi-Markets?" *Australian Economic Review*, 37, 271–286.

Bradley, S., G. Johnes, and J. Millington (2001): "The effect of competition on the efficiency of secondary schools in England," *European Journal of Operational Research*, 135, 545–568.

Buddin, R. and G. Zamarro (2009): "Teacher qualifications and student achievement in urban elementary schools," *Journal of Urban Economics*, 66, 103–115.

Burney, N. A., J. Johnes, M. Al-Enezi, and M. Al-Musallam (2011): "The efficiency of public schools: the case of Kuwait," *Education Economics*, 21, 360–379.

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

Chakraborty, K. and V. C. Blackburn (2013): “Efficiency and Equity in Funding for Government Schools in Australia,” *Australian Economic Papers*, 52, 127–142.

Chakraborty, K. and J. Poggio (2008): “Efficiency and equity in school funding: a case study for Kansas,” *International Advances in Economic Research*, 14, 228–241.

Cherchye, L., K. De Witte, E. Ooghe, and I. Nicaise (2010): “Efficiency and equity in private and public education: A nonparametric comparison,” *European Journal of Operational Research*, 202, 563–573.

Coelli, T., D. Rao, C. O’Donnell, and G. Battese (2005): *An Introduction to Efficiency and Productivity Analysis*, New York: Springer Science, 2nd ed.

Conroy, S. J. and N. M. Arguea (2008): “An estimation of technical efficiency for Florida public elementary schools,” *Economics of Education Review*, 27, 655–663.

Daraganova, G., B. Edwards, and M. Sipthorp (2013): “Using National Assessment Program–Literacy and Numeracy (NAPLAN) data in the Longitudinal Study of Australian Children (LSAC),” Tech. rep., Australian Institute of Family Studies.

Elder, T. and C. Jepsen (2014): “Are Catholic primary schools more effective than public primary schools?” *Journal of Urban Economics*, 80, 28–38.

Essid, H., P. Ouellette, and S. Vigeant (2011): “Small is not that beautiful after all: measuring the scale efficiency of Tunisian high schools using a DEA-bootstrap method,” *Applied Economics*, 45, 1109–1120.

Gonski, D., K. Boston, K. Greiner, C. Lawrence, B. Scales, and P. Tannock (2011): “Review of Funding for Schooling - Final report,” Report.

Gronberg, T. J., D. W. Jansen, and L. L. Taylor (2012): “The relative efficiency of charter schools: A cost frontier approach,” *Economics of Education Review*, 31, 302–317.

Grosskopf, S., K. J. Hayes, and L. L. Taylor (2009): “The relative efficiency of charter schools,” *Annals of Public and Cooperative Economics*, 80, 67–87.

Grosskopf, S. and C. Moutray (2001): “Evaluating performance in Chicago public high schools in the wake of decentralization,” *Economics of Education Review*, 20, 1–14.

Haelermans, C., K. De Witte, and J. L. T. Blank (2012): "On the allocation of resources for secondary schools," *Economics of Education Review*, 31, 575–586.

Haelermans, C. and J. Ruggiero (2013): "Estimating technical and allocative efficiency in the public sector: A nonparametric analysis of Dutch schools," *European Journal of Operational Research*, 227, 174–181.

Hanushek, E. A., J. F. Kain, S. G. Rivkin, and G. F. Branch (2007): "Charter school quality and parental decision making with school choice," *Journal of Public Economics*, 91, 823–848.

Hanushek, E. A. and L. L. Taylor (1990): "Alternative assessments of the performance of schools: measurement of state variations in achievement," *Journal of Human Resources*, 179–201.

Haug, A. A. and V. C. Blackburn (2013a): "Efficiency Aspects of Government Secondary School Finances in New South Wales: Results from a Two-Stage Double-Bootstrap DEA at the School Level," .

——— (2013b): "Efficiency Aspects of Government Secondary School Finances in New South Wales: Results from a Two-Stage Double-Bootstrap DEA at the School Level," .

Jepsen, C. (2003): "The Effectiveness of Catholic Primary Schooling," *The Journal of Human Resources*, 38, 928–941.

Kirjavainen, T. (2011): "Efficiency of Finnish general upper secondary schools: an application of stochastic frontier analysis with panel data," *Education Economics*, 20, 343–364.

Lamb, S., R. Rumberger, D. Jesson, and R. Teese (2004): "School performance in Australia: results from analyses of school effectiveness," Tech. rep., Centre for Post-compulsory Education and Lifelong Learning.

Leigh, A. (2010): "Estimating teacher effectiveness from two-year changes in students' test scores," *Economics of Education Review*, 29, 480–488.

Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. J. Pallen (2012): "Performance comparison of benchtop high-throughput sequencing platforms," *Nature biotechnology*, 30, 434–439.

Malmquist, S. (1953): "Index Numbers and Indifference Surfaces," *Trabajos de Estatistica*, 4, 209–242.

Mancebón, M.-J., J. Calero, A. Choi, and D. P. Ximénez-de Embún (2012): "The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006," *Journal of the Operational Research Society*, 63, 1516–1533.

Mante, B. and G. O'Brien (2002): "Efficiency measurement of Australian public sector organisations: The case of state secondary schools in Victoria," *Journal of Educational Administration*, 40, 274–298.

Marks, G. N. (2007): "Do Schools Matter for Early School Leaving? Individual and School Influences in Australia," *School Effectiveness and School Improvement*, 18, 429–450.

——— (2009): "Accounting for School-sector Differences in University Entrance Performance," *The Australian Journal of Education*, 53, 19–38.

Marks, G. N., J. McMillan, and K. Hillman (2001): "Tertiary entrance performance: The role of student background and school factors," Research report number 22, Research Report Number 22, Australian Council for Educational Research (ACER).

Masters, G. N. (2012): "Measuring and Rewarding School Improvement," .

MCEERYA (2008a): "MCEETYA four-year plan 2009 – 2012: A companion document for the Melbourne Declaration on Educational Goals for Young Australians," Report, the Ministerial Council on Education, Employment, Training and Youth Affairs.

——— (2008b): "Melbourne Declaration on Educational Goals for Young Australians," Report, the Ministerial Council on Education, Employment, Training and Youth Affairs.

Miller, P. W. and D. Voon (2012): "Government Versus Non-Government Schools: A Nation-Wide Assessment Using Australian Naplan Data," *Australian Economic Papers*, 51, 147–166.

Nghiem, H. S., H. Nguyen, R. Khanam, and L. Connelly (2013): "Does school type affect academic achievement in young children? Evidence from Australia," in *Growing Up in Australia and Footprints in Time, LSAC and LSIC Research Conference*.

Ouellette, P. and V. Vierstraete (2005): "An evaluation of the efficiency of Quebec's school boards using the Data Envelopment Analysis method," *Applied Economics*, 37, 1643–1653.

R Core Team (2013): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reback, R. (2008): "Teaching to the rating: School accountability and the distribution of student achievement," *Journal of Public Economics*, 92, 1394–1415.

Ryan, C. (2013): "What is behind the decline in student achievement in Australia?" *Economics of Education Review*, 37, 226–239.

Ryan, C. and L. Watson (2010): "The Impact of School Choice on Students' University Entrance Rank Scores in Australia," in *Society of Labor Economists Conference, London , 17–19 June*.

Shephard, R. W. (1953): *Cost and production functions*, Princeton: Princeton University Press.

Simar, L. and P. W. Wilson (1998): "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models," *Management Science*, 44, 49–61.

——— (2007): "Estimation and inference in two-stage, semi-parametric models of production processes," *Journal of Econometrics*, 136, 31–64.

Street, A. (2003): "How much confidence should we place in efficiency estimates?" *Health Economics*, 12, 895–907.

Vella, F. (1999): "Do Catholic Schools Make a Difference? Evidence from Australia," *The Journal of Human Resources*, 34, 208–224.

Wilson, P. (2008): "FEAR 1.0: A software package for frontier efficiency analysis with R," *Socio-Economic Planning Sciences*, 42, 247–254.

Zoghbi, A. C., F. Rocha, and E. Mattos (2013): "Education production efficiency: Evidence from Brazilian universities," *Economic Modelling*, 31, 94–103.