

January 2010

Measuring What Matters: Student Progress

Dr. Ben Jensen



Founding Members



Australian Government



Grattan Institute Report No. 2010-1 Reissue*, January 2010

This report was written by Dr. Ben Jensen, Program Director, School Education, Grattan Institute. All recommendations and errors contained in this report belong to the author and not Grattan Institute. Extensive research assistance was provided by Katherine Molyneux and Amelie Hunter who made a substantial contribution to this report.

We would also like to thank the members of Grattan Institute's School Education Reference Group for their helpful comments.

Grattan Institute is an independent think-tank focused on Australian public policy. Our work is thoughtful, evidence-based, and non-aligned. We aim to improve policy outcomes by engaging with both decision-makers and the community.

The opinions in this report are those of the author and do not necessarily represent the views of Grattan Institute's founding members, affiliates, individual board members or reference group members.

For further information on Grattan Institute's School Education program please go to:

<http://www.grattan.edu.au/programs/education.html>

To join our mailing list please go to:

<http://www.grattan.edu.au/signup.html>

ISBN: 978-1-925015-01-0

Table of Contents

Executive Summary.....	4
1. Introduction.....	6
2. Student progress in Australia: the need for accurate performance measures.....	10
3. School performance measures in Australia.....	13
4. Value-added measures of school performance: A more accurate and fairer measure.....	16
5. Improving school education with value-added measures of school performance.....	21
6. Using value-added measures to improve accountability.....	27
7. Promoting school choice.....	32
8. Implementing issues.....	33
9. Conclusion.....	41
10. References.....	42

Executive Summary

In Australia, a large percentage of students only reach minimum standards of literacy and numeracy. These students are spread throughout Australia's school education systems; there are few schools that do not have poor performing students who would benefit from improved education. Despite decades of increasing expenditure, student performance has stagnated (Thomson & De Bortoli 2008). We have a moral imperative to improve the performance of the 30% of year 9 students who have progressed to only the very basic elements of writing literacy.

Accurate measures of school performance are vital to improvement. The measures need to focus on student progress so that schools and teachers can focus on improving all students – particularly those most in need.

The National Assessment Program – Literacy and Numeracy (NAPLAN) assesses students' literacy and numeracy skills and is an important advance in addressing poor performance. The Federal Government's 'My School' website¹ publishes school performance scores for each school as the average of their students' NAPLAN results, comparing them to the results of 'like' schools (based on proxies of students' socio-economic background).

The publication of school performance measures is a significant step forward in achieving transparency and lifting standards in the Australian education system. However, the school performance measures published on the 'My School' website are prone to mismeasurement and may be biased against schools serving lower socio-economic communities (Ballou et al. 2004; Ladd & Walsh 2002; Raudenbush 2004; Raudenbush & Willms 1995).

Value-added scores consistently measure school performance more accurately, because they are better able to isolate the performance of schools from other factors that affect student performance. This creates a fairer system that is not biased against schools serving more disadvantaged communities. For these reasons, teachers, school associations and education unions in other countries have advocated for the introduction of value-added measures of school performance (Doran & Izumi 2004; Fitz-Gibbon 1997; Jakubowski 2008; NASBE 2005; OECD, 2008; Raudenbush 2004).

School value-added scores are calculated by comparing the progress made by each student between assessments, measuring the contribution the school makes to that progress, controlling for students' background. A school's contribution to student progress would be measured between NAPLAN assessments of literacy and numeracy at years 3, 5, 7, and 9, and students' grades in the final year of secondary school.

Value-added measures of school performance shift the focus to the student – they focus on how students learn and progress. Significant improvements come from building individualised instruction and lesson plans around multiple assessments that identify each student's learning trajectory.

School principals need to be able to identify for which students, in which subject areas and in which grade levels their school is effectively contributing to student progress. Effective programs and instruction can be expanded and less effective areas developed.

1. See www.myschool.edu.au.

In this report it is advocated that:

- The current measures of school performance published on the 'My School' website should be replaced with value-added measures of school performance because:
 - Their greater accuracy creates a fairer system, particularly for schools in lower socio-economic communities;
 - A focus on student progress rather than performance at a single point in time serves a variety of policy objectives and is more effective in improving instruction and school education.
- School principals and teachers should be empowered to use value-added measures to improve instruction and school programs. To achieve this:
 - A user-friendly information technology system should be developed that allows school principals and teachers to better analyse and then act upon their own performance data;
 - Education and training to incorporate performance assessment into instruction and school programs should be provided;
 - Resources should be provided for teachers and schools to develop programs based on value-added measures and disseminate best practice.
- Value-added measures of school performance should become an important benchmark in school evaluation. School evaluators should make their qualitative judgements of good practice in the context of value-added performance measures;
- Value-added measures of student progress should be the basis for categorising schools as under-performing. Developmental steps should be explicit, with additional support for under-performing schools; and
- School principals should be granted autonomy to effectively lead the school for which they are being held accountable. Individual teachers have continually been shown to have the greatest impact upon student performance and school principals should be empowered to determine who teaches in their school.

1. Introduction

Accurate measures of school performance are needed to address what matters in school education: student progress. Three issues highlight this need:

- A large percentage of students have only progressed to or below minimum levels of literacy and numeracy. For example, 30% of year 9 students perform at only the basic minimum levels of writing literacy (MCEETYA 2009a);
- More schools are failing to lift the performance of at least some of their students over time. Relative to other countries, Australia has wide inequality in student performance within schools and relatively less inequality between schools (OECD 2007); and
- Increases in education expenditure have not been matched by improvements in student performance. Funding in the Australian school education sector increased by 41% between 1995 and 2006 (OECD 2007). However, between 2000 and 2006, Australian student performance has stagnated in mathematics and significantly declined in reading (Thomson & De Bortoli 2008).

Additional investments in schools and education programs are not improving students' education, particularly those most in need. We have a moral imperative to address differences in student progress.

The development of school performance measures in Australia builds on the introduction of the National Assessment Program – Literacy and Numeracy (NAPLAN). NAPLAN should be strongly supported given the evidence from both Australia and other countries that high-performing education systems have more regular student assessments; efforts to turn around under-performing schools are enhanced with standardised testing like NAPLAN; and student learning improves when individualised instruction builds on assessments of student progress (Hanushek & Woessmann, 2007; Woessmann et al., 2009).

The Federal Government has driven reform for greater transparency in schools on the basis that “under-performance and failure will no longer be hidden and tolerated” (Gillard 2009a).

What is NAPLAN?

The National Assessment Program – Literacy and Numeracy (NAPLAN) is a series of assessments of Australian students in Years 3, 5, 7 and 9. NAPLAN is designed to test the requirements for literacy and numeracy common amongst the curricula of each state and territory. NAPLAN includes five assessments: reading, writing, spelling, grammar and numeracy. This year will be the third year of NAPLAN testing: the first year in which students take the assessment for a second time.

As a result, the *'My School'* website publishes school NAPLAN scores as a measure of school performance. Each school's average NAPLAN score is compared to other schools in 'like school groups'. These groups are based on census data that proxy the socio-economic background of students (DEEWR 2009a).

The *'My School'* website is considerably better than having no information published on school performance. It is a significant step forward to inform school choice and improve school accountability and school improvement initiatives. However, problems still exist because the school performance measures published on the *'My School'* website may not be accurate. Comparisons of schools' average test scores, even within like school groups, are prone to mismeasurement and can produce biased results for schools in low socio-economic areas (McCaffrey et al. 2004; McCaffrey et al. 2005; OECD 2008; Raudenbush 2004).

We need accurate measures of school performance. If not, then school choice becomes misinformed. Schools are held accountable for the wrong practices and school principals and teachers base improvements on incorrect information.

Value-added scores provide more accurate measures of school performance because they better isolate the performance of schools from other factors that affect student performance (Braun 2005; Choi et al., 2005; McCaffrey et al. 2005; Ponisciak & Bryk 2005; Tekwe et al. 2004). School performance measures similar to those produced on the *'My School'* website have consistently been shown to produce biased estimates of school performance compared to value-added modelling (Doran and Izumi 2004; Fitz-Gibbon 1997; McCaffrey et al. 2003; McCaffrey et al. 2004).

What is the *'My School'* website?

The *'My School'* website publishes the NAPLAN results for each school. A School's NAPLAN score is the average score of their students in the NAPLAN assessments. The *'My School'* website reports school performance by comparing schools' NAPLAN scores within like-school groups. Like-school groups are schools grouped together based on proxies of their students' demographic and socio-economic characteristics.

What is value-added?

A school's value-added score is different to its NAPLAN score that is presented on the *'My School'* website. A school's value-added score represents the contribution the school makes to the progress of its students. This is calculated using a statistical model. The model compares the progress made by each student to the progress of other students with the same initial level of attainment, controlling for background factors.

School value-added scores measure the contribution schools make to student progress rather than the current focus on student performance at a single point in time. For example, value-added scores measure student progress between NAPLAN assessments in years 3 and 5 – the increase from a student's score in year 3 numeracy to their score in year 5 numeracy. It compares the progress made by each student at each initial level of performance in year 3, and calculates the contribution the school makes to that progress by year 5, controlling for students' background. Student progress would be measured between all NAPLAN assessments of literacy and numeracy at years 3, 5, 7, and 9 and students' grades in the final year of secondary school.

Teachers, school associations and education unions in other countries have advocated for the introduction of value-added modelling as the greater accuracy creates a fairer system, particularly for schools serving more disadvantaged communities (Jakubowski 2008; NASBE 2005; OECD 2008).

The current measures of school performance published on the *'My School'* website should be replaced with value-added measures of school performance, given their greater accuracy and fairness to schools serving different student populations. A focus on student progress rather than performance at a single point in time, is not only more accurate but serves a variety of policy objectives and is more effective in improving instruction and school education.

While accurate measures of school performance are essential for a fair system, they are not an end in themselves, but a basis for action. Governments should ensure that school value-added measures are used constructively to improve school effectiveness:

- At the student level, it complements teachers' assessments by providing an evidence base for teachers to shape instruction to meet the needs of both high and low performing students;
- At the school level, principals can use value-added measures to identify for which students, in which year levels and in which subjects they are effectively contributing to student progress. From this, school principals can identify and promote teaching practices and programs that have the greatest success in improving student progress. School principals and teachers require access to this data through a user-friendly information technology system; and
- At the system level, policy makers can identify programs that are truly effective in lifting student progress. This is particularly important given that substantial increases in funding have been unsuccessful in improving students' progress.

Real change occurs within schools and this where the focus on student progress is crucial. In the current system, the focus is on what students' NAPLAN scores were last year. Value-added measures focus on student progress and are therefore more effective for informing instruction: the focus shifts to how students learn and progress. The greatest inequality in Australian student performance occurs within schools (Thomson & De Bortoli 2008). It is therefore essential to shift the focus from the difference in performance between schools, to inequality within schools. Significant improvements come from building individualised instruction and lesson plans around multiple assessments that identify each student's learning trajectory.

Publishing school performance measures will increase accountability in school education. Both governments and schools should be held accountable for the quality of education they provide. Holding school principals accountable for the performance of their school (and having that performance made public) requires they be empowered with sufficient autonomy to act on value-added measures.

School principals should be granted autonomy to make important decisions that affect instruction. Individual teachers have continually been shown to have the greatest impact upon student performance and school principals should be empowered to determine who teaches in their school (Caldwell & Harris 2008; OECD 2005b; OECD 2009).

Opponents of the publication of school performance measures are often concerned about league tables that unfairly stigmatise schools in poorer communities. School performance measures as published on the *'My School'* website are susceptible to being converted to league tables as it is still incorrectly being assumed that schools' average NAPLAN scores are a measure of school performance (even if schools are reordered into like-school groups). Value-added scores are different and are a more accurate measure of school performance; they would more effectively counter ill-conceived league tables published in the media.

2. Student progress in Australia: the need for accurate performance measures

Value-added measures of school performance have the power to directly address what matters in school education: student progress. Three issues highlight the need for accurate measures of student progress:

- A large percentage of students at or below minimum levels of literacy and numeracy;
- Large inequality in performance within schools; and
- The disconnect between education spending and student progress.

2.1. Percentage of students at or below minimum levels of literacy and numeracy

The development of the NAPLAN assessments of literacy and numeracy has increased attention on standards and facilitated analysis of poor performing students (COAG Reform Council 2009). The COAG Reform Council in 2009 noted that the NAPLAN minimum levels of literacy and numeracy should be considered as low levels of achievement for the purposes of policy analysis and reform.

The 2009 NAPLAN Summary Report states:

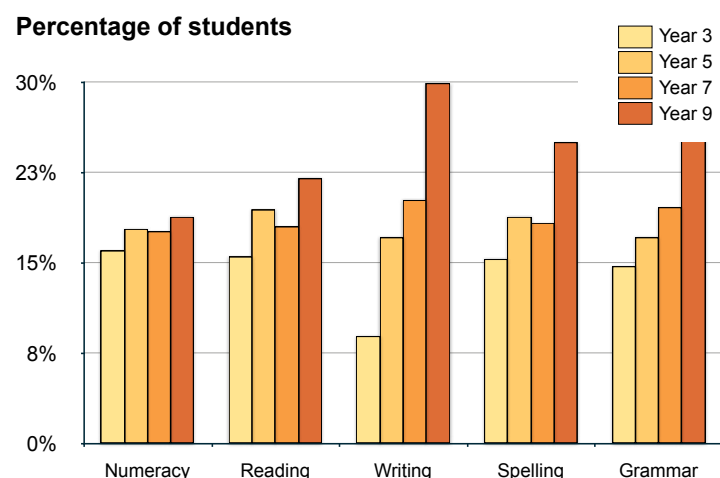
Students whose results are in the minimum standard band have typically demonstrated only the basic elements of literacy and numeracy for the year level. Students whose results are in the lowest band for the year level have not achieved the national minimum standard for that year, and need focused intervention and additional support to help them achieve the skills they require to progress in schooling.

Little comfort can be garnered from having a significant portion of students demonstrating only the basic elements of literacy and numeracy. The focus here therefore is on students performing at or below minimum standards rather than simply below the minimum. This is particularly important given the substantial social and economic difficulties encountered by students and adults who have mastered only the basics of literacy and numeracy.

Figure 1 presents the percentage of Australian students performing at or below the minimum basic standards of literacy and numeracy. A substantial percentage of students perform at or below minimum numeracy and literacy standards in Australian schools. Nearly one-third of Australian Year 9 students demonstrate, at best, the basic minimum writing literacy. Although there is variation between jurisdictions, states and territories follow similar patterns.² In addition, at least one-quarter of Australian Year 9 students perform at or below the minimum standard of spelling and grammar.

2. For example, the percentages of Year 9 students performing at minimum writing literacy range from 25% of students in Victoria to 35% in Queensland and 52% of students in the Northern Territory.

Figure 1.
Percentage of Australian students at or below
minimum standards of literacy and numeracy (2009)



Source: 2009 NAPLAN Summary Report.

Greater attention needs to be paid to the progress, or lack thereof, made by students through their school education. A snapshot of student performance at a single point in time does not provide sufficient information to address declining student progress, emphasising the need for value-added measures.

2.2. Dispersion of under-performing students within schools

The OECD Programme for International Student Assessment (PISA) assesses students in reading, mathematics and science. The skills assessed in PISA are not narrow. They concentrate on problem-solving abilities and correlate closely with higher rates of economic growth (Hanushek & Woessmann, 2007). PISA provides valuable information on student performance, including two troubling issues for Australian school education. First, inequality in performance in scientific literacy between Australian students was 11% higher than the OECD average indicating a degree of inequality that should be a concern to policy makers. Second, the vast majority of inequality in performance amongst Australian students was within schools. 81% of the total variance in students' science performance was within schools, compared to 18.4% between schools and a negligible 0.6% between states and territories (Thomson & De Bortoli 2008). The inequality in performance in Australia explained by differences within schools, rather than between schools, is the 5th highest of all OECD countries (OECD, 2007).

Greater attention needs to be paid to the progress, or lack thereof, made by students through their school education.

Inequality in student performance within schools is relatively weakly correlated with social and economic background, suggesting that within schools there are other factors at play in determining student outcomes (Thomson & De Bortoli 2008). This appears to have received little attention in education policy debate with much of the focus on inequality between schools, particularly schools in poorer neighbourhoods (MCEETYA 2008).

Given that the majority of inequality in student performance is within schools, blanket school-wide policies may not target the students that most need assistance. It is vitally important we understand what is causing this inequality between students and what programs are making a difference to individual student progress. This report argues that value-added measurement is one of the best tools to equip school principals and teachers to address the inequity in performance within schools. Analysis of the progress made by each individual student facilitates development of instruction and programs to reduce inequities.

2.3. Spending increases have not improved student performance

Funding in the Australian school education sector increased by 41% between 1995 and 2006 (OECD 2007). However, between 2000 and 2006, Australian student performance in PISA stagnated in mathematics and significantly declined in reading (Thomson & De Bortoli 2008). It is clear that resources are not being efficiently allocated. Effective programs are needed to lift the progress of the 30% of Australian year 9 students with only the basic minimum standards of writing literacy. More accurate performance measures are required to not only develop effective programs for these students, but to evaluate their impact upon student progress. It is clear that past education investments are not working for those students most in need.

3. School performance measures in Australia

The Council of Australian Governments (COAG) reached a National Education Agreement in November 2008 that emphasised the importance of high quality school information, including national reports on education outcomes and individual schools. Three objectives were emphasised by COAG Ministers in their principles for reporting on schooling in Australia agreed in 2009:

1. High quality accountability to students, parents, carers and the community;
2. Tracking the achievement of the COAG targets; and
3. Providing evidence to support future policy reforms and system improvements including the aim of better directed resources (MCEETYA 2009b).

These objectives can be categorised under the headings of school choice, school accountability and school improvement. To meet these policy objectives, the Federal Government has begun publishing NAPLAN results as a measure of school performance on the *'My School'* website.

3.1. National Assessment Program – Literacy and Numeracy

NAPLAN is a series of annual assessments of Australian students in years 3, 5, 7 and 9, designed to assess literacy and numeracy, the facets of which are common amongst the curricula of each state and territory.³ NAPLAN includes five assessments: reading, writing, spelling, grammar and numeracy.

NAPLAN results are standardised every year in order that the tests provide a consistent scale of achievement. For each learning domain (reading, writing, etc) results are on a single scale between 0 and 1000, with a standard deviation of 100. This scale is divided into ten National Assessment Program (NAP) achievement bands: each year level's performance is assessed within six of these NAP bands. As such, the performance of year 3 students is presented between bands 1 and 6; year 5 students are assessed between bands 3 and 8; Year 7 between 4 and 9; Year 9 between 5 and 10. For each year level, a result in the lowest band indicates student performance below minimum standards of numeracy or literacy. The second lowest band equates to the minimum standards for the year level; and the higher NAP bands indicate a result above the minimum standard of achievement required for the year (DEEWR 2009c).

The NAPLAN assessments are intended to provide schools and policy makers with a data-rich source of information about the performance of Australian schools and their students. Literacy and numeracy skills are significant features of school curricula at all year levels, rendering the assessment of these skills the ideal media for measuring school performance (DEEWR 2009c).

3. It is important to note that it is envisaged that the NAPLAN assessments will be refined once the planned national curriculum has been designed and implemented.

3.2. 'My School' website

Using NAPLAN data to assess the performance of individual schools is the most controversial element of the Federal Government's 'My School' website. The website is intended to provide parents with a wealth of information about their child's school, with NAPLAN intended to form only part of the picture. School reports include:⁴

- Mean scaled scores for each tested domain: data may be compared to the national mean and the mean of 'like schools';
- A list of 'like schools' throughout Australia. Parents are able to see detailed results of these schools, that may include as many as 60 schools;
- A breakdown of the percentage of students in each NAP band for each domain. This data is then compared with the national and 'like school' results
- A list of 20 local schools, including detailed results;
- Information about student and staff population; and
- A school result on the Index of Community Socio-Economic Advantage (ICSEA).⁵

School NAPLAN scores are presented in the 'My School' website alongside other schools in 'like school groups.' These school groups are determined by each school's ICSEA value; an index of school context designed to include those factors which impact on NAPLAN results. These include remoteness, indigenous population and proxies of the socio-economic status of the student population. Socio-economic status is measured by the socio-economic characteristics of the area in which a student lives (according to ABS census collection district (DEEWR 2009a)).

4. For brevity, we have referred to school performance measures soon to be published on the 'My School' website as 'current measures'.

5. At the time of publication, school results had not yet been released on the 'My School' website (<http://www.myschool.edu.au>). This information is collected from illustrative examples of what will be published.

Like-school groups are formed by ranking Australian schools by their ICSEA value. Each school's like-school group consists of up to 60 schools adjacent to them on the ICSEA ranking.⁶ School performance is presented by comparing the NAPLAN scores of a school's students to the NAPLAN scores of 'like schools' (DEEWR 2009b).

The 'My School' website is considerably better than having no information published on school performance. However, comparisons of schools' raw test scores, even within like school groups, have been shown to be prone to mismeasurement and can produce biased results for schools in low socio-economic areas (McCaffrey et al. 2004; McCaffrey et al. 2005; OECD 2008a; Raudenbush 2004).

This is because an array of student background socio-economic characteristics influences student performance. If the full effect of all of these characteristics is not fully captured, school performance measures become biased. As this is most important for students from lower socio-economic backgrounds, the bias is greatest for the schools educating these students.

But perhaps of most importance is the evidence showing that even the most complex modelling of student background cannot match the accuracy of value added measures (Hægeland & Kirkebøen 2008).

Comparisons of schools' raw test scores, even within like school groups, have been shown to be prone to mismeasurement and can produce biased results for schools in lower socio-economic areas.

6. For some schools, like-school groups may contain considerably less than 60 schools.

4. Value-added measures of school performance: A more accurate and fairer measure

School value-added scores measure the contribution schools make to student progress. For example, a value-added score can measure student progress between NAPLAN assessments in years 3 and 5. It compares the progress made by each student at each initial level of performance in year 3, and calculates the contribution the school makes to that progress by year 5, controlling for students' background. Student progress can be measured between all NAPLAN assessments of literacy and numeracy at years 3, 5, 7, and 9, and students' grades in the final year of secondary school.

Value-added measures are better than raw test scores or comparisons in like-school groups assessing the contributions made by individual schools. They are more accurate, less biased against schools in lower socio-economic communities, less vulnerable to data collection issues, and more useful for improving schools and teaching.

4.1 Measuring schools' value-added scores

Value-added measures focus on student progress rather than performance at a single point in time as is the case with current measures published on the 'My School' website. The OECD defined school value-added as:

The contribution of a school to students' progress towards stated or prescribed education objectives (e.g. cognitive achievement). The contribution is net of other factors that contribute to students' educational progress (OECD 2008).

Given the above definition, value-added modelling was defined by the OECD as:

A class of statistical models that estimate the contributions of schools to student progress in stated or prescribed education objectives (e.g. cognitive achievement) measured at at least two points in time (OECD 2008).

The key feature is the focus on student progress between "at least two points in time."⁷ School value-added measures derive from applying a statistical model to a dataset that should contain student and school identifiers, a range of student demographic and socio-economic background information, and at least two successive student assessment scores.⁸ A school value-added score represents that part of student performance that cannot be explained by other student background variables included in the model.

7. It is important to realise that this definition of value-added modelling differs from at least the terminology used by the Victorian Government (DEECD 2009). The intake adjusted measures developed by the Victorian Government are sometimes referred to as value-added measures even though they do not measure student progress and only measure student performance at a single point in time.
8. Socio-economic factors may include a student's: age; gender; indigenous status; whether or not they are new to a school; family education and occupation; country of birth and migration background; language preferences; grade repetition; and student learning difficulties. For a more in depth discussion, please refer to Section 8.6.

Similar to residuals in a standard regression model, these scores sum to zero which means that a school with a value-added score of zero has students that have progressed at the average growth rate between the two assessments.⁹ In this sense, value-added measures are relative rather than absolute performance measures, differentiating them from many statistical models of school performance. Students have still progressed in a school with a value-added score of zero, but have progressed at the average rate for students with the same initial attainment level, controlling for differences in student background. Students in a school with a positive value-added score have progressed at a rate greater than the average. Standard errors are also produced with each estimate of school value-added, enabling a categorisation to be made of schools that are performing statistically significantly above or below the average.

As at January 2010, the NAPLAN assessments have not been in place for a sufficient period to allow individual students to sit two NAPLAN assessments (for example, a student who sat the Year 7 assessment in 2008 will sit his or her second NAPLAN assessment in Year 9 of their school education in May 2010). However, with the assessments undertaken by Australian students this year, value-added measures of school performance can be presented after May 2010.

4.2. Benefits of value added measures of school performance

Value-added modelling is an advance in the measurement of school performance, producing more accurate measures and reducing the problems that plague other methods (Braun et al. 2008; Fitz-Gibbon 1997; Leigh 2009; Raudenbush 2004). The key reason that value-added measures are more accurate is that they are better able to isolate the performance of schools from other factors that affect student performance (Braun 2005; Choi et al. 2005; McCaffrey et al. 2005; Ponisciak & Bryk 2005; Tekwe et al. 2004).

Unadjusted raw test scores provide little, if any evidence of school performance. Such scores better reflect the socio-economic status of a school's students than the performance of the school (Bethell 2005). Unfortunately, schools are often ranked in the media based upon students' raw scores. Lists of schools, or league tables, are produced based on the average score of each school's students in their final year of secondary school.

Value-added measures provide more accurate measures of school performance.... they are better able to isolate the performance of schools from other factors that affect student performance.

9. We recommend that the scale of the value-added scores be altered so that 100 is the mean score rather than zero. For a further explanation, please refer to Section 8.5.

Headlines such as “Tassie High School Rankings” in The Hobart Mercury, ‘Your School’s VCE Performance’ in The Age, and ‘The Top 100’ in the Sydney Morning Herald rank schools purely on their students’ raw scores that have little relation to actual school performance. Articles such as these occur in most Australian states and territories and rely on rankings that are inaccurate and unfair to schools serving low-socio-economic status communities (Perkins 2009; Patty & Gilmore 2009).

The school performance measures published on the ‘*My School*’ website cannot properly counter these league tables as it publishes the same NAPLAN scores but simply groups them differently (into like-school groups).

Value-added measures of school performance are therefore required to counter ill-conceived media reporting. A focus on value-added measures does not rely on simply re-ordering schools’ NAPLAN scores into like-school groups. There are benefits to continuing to publish school’s NAPLAN scores as there are substantial benefits to having this information. Value-added scores should be published alongside NAPLAN scores, emphasising that value-added scores are the ‘true’ measure of school performance. This provides a clear argument to counter ill-conceived league tables that are based solely on NAPLAN scores.

The ‘*My School*’ website compares schools’ NAPLAN scores within like-school groups. This approach includes student background characteristics which is an improvement upon league tables that focus only on NAPLAN scores.¹⁰ Yet, this is not optimal if the objective is to isolate the impact of schools upon student performance and thus measure school performance (Braun et al. 2008; Hægeland & Kirkebøen 2008; NASBE 2005).

These types of modelling have consistently been shown to produce biased estimates of school performance compared to value-added modelling (Doran & Izumi 2004; Fitz-Gibbon 1997; McCaffrey et al. 2003; McCaffrey et al. 2004). This is particularly so if the student background characteristics included in the estimates of school performance are incomplete or endogenous (Hægeland & Kirkebøen 2008).

The quality of student background data in Australia that is currently used to measure school performance lacks the breadth and accuracy to properly isolate school performance (this is further discussed in Section 8 which argues for more student-level data to be collected). The like-school groups calculated by the Federal Government rely on census data for where a student resides to control for differences in students’ background that would influence their performance in NAPLAN assessments. This is a crude measure of student background as the census data depends on where a student lives rather than concentrating on differences between individual students. This could increase the bias in the current school performance measures published on the ‘*My School*’ website, particularly given the relatively large inequality in student performance within schools (OECD, 2007). Value-added measures are better equipped to deal with these problems as the focus on student progress reduces the impact of problems with student background data (Raudenbush 2004).¹¹

10. This is also the approach of the Victorian Government’s intake adjusted school performance measures and reflect the cross-sectional models used in a number of other countries (Aitken & Longford, 1986; DEECD, 2009; Willms & Raudenbush, 1989).

11. Nevertheless, this report argues strongly for the need to collect better student background data, which is more fully discussed in Section 8.6.

Research comparing value-added modelling with other types of statistical modelling in Norway underscores this point. Hægeland and Kirkebøen (2008) show that the extent to which students' socio-economic background characteristics are included in statistical modelling (that is similar to the methods used in Australia) greatly affects findings of school performance. Norway has incredibly rich data sources to properly control for student background that far exceed the breadth and depth of data available in Australia. This should create more accurate measures of school performance. However, the authors found that even the best model with the most extensive student background characteristics produced poor measures of school performance compared with a value-added model. Even a simple value-added model with little student background data had greater explanatory power (almost double) than more complex modelling that did not measure student progress.

Further studies have shown that no matter how well a model controls for student background characteristics, if they fail to measure student progress, unobserved background characteristics can create biased school performance measures, particularly in systems (such as in Australia) that serve diverse student populations (Ballou et al. 2004; Ladd & Walsh 2002; McCaffrey et al. 2004; McCaffrey et al. 2005; Raudenbush 2004).

The greater accuracy and applicability to school improvement initiatives of value-added measures have led other countries to adopt such methods (OECD 2008). Fitz-Gibbon (1997) reviewed the various methods available to the British Government to measure school performance and favoured value-added modelling of student progress, concluding it was statistically valid and readily understood. The greater accuracy resulted in head teachers in England favouring the introduction of a system of value-added modelling (Fitz-Gibbon 1997).

In other European countries, further analyses have shown that value-added modelling is preferred due to its greater accuracy and more effective use within education systems (Ryska 2006; van de Grift 2007). Institutions such as teacher unions and school associations in a number of countries have also supported the introduction of value-added modelling as the greater accuracy creates a fairer system, particularly for schools serving more disadvantaged communities (Jakubowski 2008; NASBE 2005; OECD 2008).

This may be particularly important given the calls in Australia that the current system is unjust and produces inaccurate measures of school performance (AEU 2009; Ferrari, 2009b; Tomazin 2009).

Value-added measures ... create a fairer system, particularly for schools serving more disadvantaged communities.

It must be said that these calls have also included some criticisms of value-added measures that have focused on their interpretation and applicability to the current context (Masters et al., 2008). In particular, it has been argued that value-added scores are difficult to explain and of secondary interest to parents and other groups (Masters et al. 2008). This is in contrast to the successful implementation of value-added in a number of countries that acknowledges the need to explain value-added scores to parents and other stakeholders, but recognises that this is a small price to pay for more accurate measures of school performance. This research seems to have weighed the benefits of value-added scores against school NAPLAN scores (Masters et al. 2008). Instead, value-added scores should be published alongside schools' NAPLAN scores as this information is valuable to all stakeholders. Both should be published but value-added scores should be considered the 'true' measure of school performance given its greater accuracy (Ballou et al. 2004; McCaffrey et al. 2004; McCaffrey et al. 2005; OECD, 2008; Raudenbush & Willms 1995).

Some critics of value-added have cited evidence of the inappropriate use of value-added scores in debate in other countries (Donnelly 2010). This has often focused on inappropriate teacher pay-for-performance schemes (National Research Council, 2010). Such initiatives are not suggested here. Multiple sources of information need to be used in any evaluative process.

Policy recommendation:

The current measures of school performance published on the 'My School' website should be replaced with value-added measures of school performance, given their greater accuracy and fairness to schools serving poorer communities. A focus on student progress rather than performance at a single point in time is not only more accurate, but also serves a variety of policy objectives and is more effective in improving instruction and school education.

5. Improving school education with value-added measures of school performance

Creating the most accurate measures of school performance is not an end in itself, rather it is a basis for action. A key advantage of value-added measures is that they provide useful data for teachers, school principals and policy makers to inform evidence-based decision-making and improve instruction. To this end, the data must be made available and accessible to schools and teachers: it is at the classroom level where change will have the greatest impact on student progress (Dudley 1999).

Unfortunately, in some education systems standardised testing led to few changes in school level practice (OECD 2005a). Schools did not use the data to improve instruction and policy makers failed to address the complex organisational change required to institute school-level data-based decision-making (O'Day 2002). However, value-added measures of student progress are more readily incorporated into the work of school principals and teachers.

A number of education systems have implemented initiatives to facilitate effective use of performance data within schools (Dudley 1999). Education and training programs have been successfully implemented (e.g. Griffin et al. forthcoming; Jakubowski 2008), stakeholder and communication strategies developed and executed (e.g. Ray 2006), and considerable investments have been made to enable schools to analyse and then respond to their own performance data to improve the education offered to their students.

While the 'My School' website is a Federal Government initiative, it is important to recognise the multiple education systems operating within Australia between and within states and territories. The recommendations made here will therefore require changes of different magnitudes in each system. Nevertheless, value-added measures are more useful to improving school performance and it is important to illustrate how they should be used effectively within all schools.

5.1. Using value-added measures to improve instruction

There is growing evidence of the benefits of data-based decision-making within schools (Black and William 1998; EPPI 2002; Griffin et al. forthcoming; OECD 2008, 2005a; Taylor et al. 2005). Value-added measures of student progress provide accurate information that allows schools to direct resources, gauge the impact of various school programs, and take steps to improve school effectiveness (Braun 2005; Ray 2006).

Developing education programs and shaping instruction based on the progress made by individual students have been effective in lifting the performance of both low and high achieving students (Griffin 1990; Halverson et al. 2005; Klieme et al. 2006; Mokhtari et al. 2007; Phillips et al. 2004; William et al. 2004). Formative assessment and scaffolding of instruction utilise an evidence base that is created through documented teaching, learning and assessment within schools to create individualised instruction that feeds and responds to each students' progress and learning needs (Vygotsky 1986).

The emphasis on student progress highlights the need to move away from summative assessments that summarise students' attainment up to a given point in time. Instead, it highlights individual progress and the next stage in students' learning trajectories, upon which effective instruction is developed. This requires a shift in mindset from what students have achieved to how student learning will progress.

Effective formative assessment and scaffolding of instruction is often not a core element of teachers' education and training. In fact, the important links between different instructional practices and student achievement is not always made by teachers (Timperley & Robinson 2001). Changes to teacher education and training need to be made to ensure that teachers can successfully incorporate value-added and other assessment information into effective instruction.

5.2. An information technology system for school principals and teachers

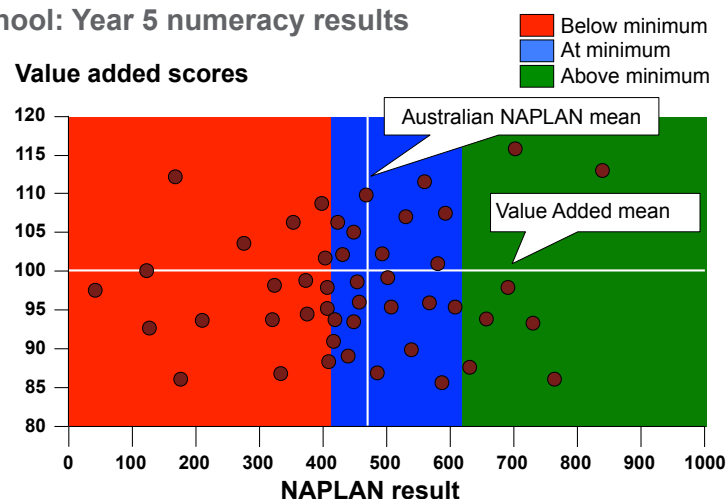
An effective information technology system should be implemented with a user-friendly interface that empowers school principals, teachers and other educators to analyse their school's performance data. These operate in a number of countries but are most prominent in England and in various states in the USA (OECD 2008). These systems help school principals and teachers analyse the impact they have on their students and to learn about the effectiveness of different aspects of their school. Two examples of how this could operate in Australia are provided below.

Figure 2 presents Year 5 numeracy results for a hypothetical 'city school' including both raw NAPLAN scores and value-added scores. City School has students who performed below the Australian average in the Year 5 NAPLAN numeracy assessment. Their average score was 469 compared to the Australian average of 487 (depicted with a white vertical line). The Figure highlights the students in each performance band with the blue band equal to the minimum standard and those in the red sector performing below the minimum numeracy standard. On the vertical axis is the value-added score of what the school has contributed to each student's progress. The brown dots represent the NAPLAN and value-added scores of each student. It may assist interpretation if we divide this Figure into quarters:

Shaping instruction based on the progress made by individual students has been effective in lifting the performance of both low and high achieving students.

- The upper left hand quarter has students who performed below (or close to) the NAPLAN average and for whom the school is contributing additional value to their progress (above the value-added average of 100)
- The upper right hand quarter has students who performed above (or close to) the Australian average and for whom the school is contributing additional value to their progress
- The lower right hand quarter has students who performed above (or close to) the Australian average but for whom the school is contributing less than average value to their progress
- The lower left hand quarter has students who performed below (or close to) the Australian average and for whom the school is contributing less than average value to their progress

Figure 2.
City school: Year 5 numeracy results



What makes this Figure particularly useful is that school principals and teachers can analyse the contribution they are making to their students' performance.¹² For example, consider the students in the upper left-hand quadrant of the figure. These students have poor NAPLAN scores but the school is effective in increasing the progress these students are making. Clearly, analysis of only NAPLAN scores would show that all students with a low NAPLAN score are poorly performing, however, this fails to distinguish between students who are consistently low performing and students who are progressing well from an initial low attainment base. Without this distinction schools cannot learn about which programs are helping students improve. The students are simply all classed as having low numeracy achievement and the school is considered under-performing for all of these students. When 'controlling' for student background by comparing like-school groups (or the Victorian intake adjusted measures), the school principal and teachers at City School may be able to point to other like schools and show that they have similar students performing at low levels. But they would learn nothing about where their school is and is not making progress.

12. An effective interface would also enable educators to 'click' on any of the students to learn more about their characteristics and their performance in other areas (including school assessments of their abilities).

Table 1 provides an example this time of the value-added profile for the hypothetical Lakeside Park Secondary School. It gives the school principal (Mrs Cleveland) a great deal of information on which subject areas and year levels should be the focus of more attention and resources.

Conversely, Mrs Cleveland can also identify areas which have been successful in lifting student progress that could then provide the basis for programs across the school.

Although not shown in this illustrative table, Lakeside Park's students tend to score relatively well on the NAPLAN assessments, but the NAPLAN numeracy scores in later years of secondary school are lower than would be hoped. Mrs Cleveland was concerned but knew that NAPLAN scores alone did not reflect the success of her teachers.

Instead, she chose to study the value-added profile shown in Table 1. In this example, value-added results that are statistically significantly above the average are highlighted in green and results statistically significantly below the average are highlighted in red. The figures in the Table represent the average contribution of the school to student progress compared to students from other schools with the same initial attainment, controlling for socio-economic background. For simplicity, this figure can be converted to a percentage or a 'gain score'. Column 2 would then show, for example, that in 2012, Lakeside Park increased their Year 9 students' writing scores by 15 score points compared to other schools, with the same initial attainment, controlling for socio-economic background. This enabled Mrs Cleveland to easily identify trends within her school.

Table 1.
Value-added performance results: Lakeside Park Secondary School

Value-Added Results: 2012 – 2014					
		Year 8	Year 9	Year 10	Year 12
2014	Reading	101	104	109	112
	Writing	102	108	102	101
	Spelling and Grammar	99	107	101	103
	Numeracy	88	87	98	100
2013	Reading	99	100	111	99
	Writing	109	106	102	104
	Spelling and Grammar	103	110	100	102
	Numeracy	89	86	98	97
2012	Reading	97	102	98	97
	Writing	113	115	97	99
	Spelling and Grammar	99	109	101	100
	Numeracy	82	93	99	100

Note: The final column 'Year 12' corresponds to grades achieved by students in their final year of secondary school.

Mrs Cleveland noted that even though the raw numeracy scores of her upper-secondary school students were lower than hoped, value-added scores showed that the cause of this decline was actually in the lower-secondary school.

The NAPLAN numeracy scores showed that decline occurred between Year 7 and 8 and 8 and 9. Mathematics teachers in the lower secondary school were asked to evaluate their program and any significant teaching practices that may have caused the reduction in student progress. Further investigation showed that the recently introduced practice of dividing mathematics classes according to aptitude was having negative results. Value-added analysis showed that this had led to a marked deterioration in attainment, particularly amongst less able students, who were no longer able to benefit from the higher performing students in their classroom. Mrs Cleveland has therefore abandoned the policy. Simultaneously, she has launched a development program for less able students, with the hope of correcting the deterioration in attainment amongst these students.

In 2012, Mrs Cleveland also noted the excellent value-added results in literacy amongst year 8 and year 9 students. Evaluation of a new cross-grade teaching program developed by year 7 and 8 teachers led to the program being expanded and implemented across the school, and has produced a steady improvement in value-added results for literacy amongst Lakeside Park's students.

These examples provide simple illustrations of how value-added measures can be effectively used to lift school and student performance. In contrast, NAPLAN scores produced misleading and less useful results. Both sets of data should be supplemented at the school-level with teacher and school assessments to provide an operable portfolio of assessments for each student.

Value-added measures allow schools to direct resources, gauge the impact of various school programs, and take steps to improve school effectiveness.

5.3. Developing and sharing effective performance

Resources should be provided for teachers and schools to better analyse value-added measures, develop programs and disseminate best practice. Programs could focus on a range of issues such as: formative assessment and scaffolding of instruction; programs to lift the performance of particular groups of students; remedial classes to improve low performers; peer critical learning programs; or school wide initiatives such as mentoring and additional teacher training, increased professional collaboration and cross-grade and classroom teaching. The focus of the program may matter less than the methodology which would require school principals and teachers to use value-added performance measures as a component of the development of their programs and then importantly, for monitoring and evaluation.

This would encourage the use of performance data so that it has a greater impact at the school-level. It would recognise school principals and teachers as professionals who have the greatest impact on student learning. School principals and teachers who have developed successful programs could be rewarded through inclusion in specifically developed peer reviewed journals and conferences. Positions could be created whereby teachers and school principals with the most effective programs work with other schools to spread best practice.

Currently, Australia's teachers believe they are not recognised for being innovative in their schools and classrooms.¹³ Their efforts are not recognised nor rewarded and this should be considered a serious problem for those that consider schools as innovative centres of learning (OECD 2009b).

Recognition of innovative practices evaluated with value-added measures would send powerful signals about the skills and capabilities of school principals and teachers. Schools could also publicise their successful programs, thereby promoting school choice in their community.

Policy recommendation:

Resources should be provided for teachers and school principals to better analyse value-added measures, develop programs and disseminate best practice.

School principals and teachers should be empowered to use value-added measures to improve instruction and school programs. To achieve this:

- A user-friendly information technology system should be developed that allows school principals and teachers to better analyse and then act upon their own performance data; and
- Education and training for school principals and teachers to incorporate value-added measures and other assessment information into instruction and school programs should be provided.

13. This issue will be further explored in a forthcoming Grattan Institute education report focusing upon the development and evaluation of Australian teachers.

6. Using value-added measures to improve accountability

Accountability in Australian school education has historically focused upon the adherence to regulations and the evaluation of school processes, normally by a school evaluator, and the level of inputs in education. The Federal Government appears intent on changing this focus to one on outcomes (Gillard 2009a), and this is being followed by states and territories (MCEETYA 2009).

Publishing school performance measures on the 'My School' website will increase school accountability. Publicly releasing performance data holds school principals and teachers to account for the education they provide to students and their use of public funds (Ladd & Figlio 2008). However, if school performance is incorrectly measured, the system loses credibility and good teachers and school principals are unfairly and publicly labelled as poor.

Within schools, the focus on literacy and numeracy to measure the performance of schools provides an incentive to focus on these areas of education. A school principal could concentrate more of his or her school's resources on subjects and students that bring the greatest gain in the NAPLAN assessments (Dixit 2002).

Additional class time or instruction may be scheduled or more resources (e.g. additional teachers) devoted to existing class time to increase the school's NAPLAN score. However, the evidence of this occurring in other systems is mixed (Jacob 2002; O'Day 2002). Additionally, this practice should not be viewed as purely negative. There may be benefits to focusing on specific subject areas or on specific students, particularly given the large proportion of Australian students falling to or below minimum literacy and numeracy levels.

An often forgotten aspect of publishing school performance measures is increased political accountability (Ladd 2007). Governments should be held responsible for improving school performance and the public will now be able to track the impact made by Governments and their policies and programs (COAG Reform Council 2009). In Australia, this is particularly important given the comparisons that can be made between states and territories. Yet, if political accountability for school education is to be fair and helpful in informing decision-making then the measures of school performance must be accurate.

If school performance is incorrectly measured, the system loses credibility and good teachers and school principals are unfairly and publicly labelled as poor.

6.1. School evaluations

Accurate and informative school evaluation feeds all three of the policy objectives discussed in this report: school improvement, school accountability and school choice. Value-added measures of school performance improve the effectiveness and accuracy of school evaluations and also the efficiency of the evaluation process itself.

Accurate measures of school performance can guide an evaluation of a school by highlighting the areas operating more and less effectively. The time available for the evaluation can then be concentrated on those areas most in need while learning from more effective areas of the school. This increases both efficiency and effectiveness through greater analysis of inputs, processes and their impact upon performance.

Evaluation of processes is subjective by nature with teachers and school principals assessed against what is considered best practice. Yet, it is difficult to determine best practice without an accurate measure of performance. The development of value-added modelling in education systems in OECD countries has led some school inspectorates and evaluators to undertake reform to ascertain the extent to which their (often subjective) assessments of best practice were actually having the desired impact upon students. For example, the Dutch inspectorate re-evaluated their qualitative assessments of best practice once they had gained value-added performance data.

They identified key organisational and process standards that influence teaching and learning within schools. These standards were successful in turning around under-performing schools and included:

- School and educational climate;
- Adherence to the curriculum;
- Effective learning time;
- Structured and engaging instruction;
- Remedial measures for low-performing students;
- Strong support for teachers from the school principal;
- Effective teacher cooperation;
- Regular student assessments, particularly with standardised testing; and
- Monitoring of the quality of instruction.

With the development of value-added measures they were able to identify how performance in these standards influenced student progress in schools (van de Grift & Houtveen 2006). Previously, inspectors had been using less reliable information to assess standards (OECD 2008).

The entire evaluative process is improved if evaluators have accurate measures of performance at their disposal to complement other quantitative and qualitative measures to create a more complete picture of school performance.

There is little value in simply naming and shaming under-performing schools.

6.2. Actions for under-performing schools

There is little value in simply naming and shaming under-performing schools.¹⁴ Instead, transparent and explicit steps need to be taken to address under-performance. This is an important area that needs greater attention than the scope of this report. However, an illustrative example is provided below that builds on experience in other countries that have identified key steps to turn around under-performing schools. These steps may include:

Year 1 – An external evaluation is triggered when a school is identified as under-performing (according to their value-added score). This would be the first step in a developmental plan with measurable milestones which aim to increase school performance. An external evaluator would work with the school to improve the internal evaluative capabilities and determine the key issues in the school causing under-performance. The areas identified by the Dutch Inspectorate in the previous example would be particularly important.

Year 2 – If under-performance persists in the following year then the school principal would be given 2 years to lift the school to an adequate level of performance. This would also entail an extensive external evaluation of all staff with individual development programs put in place and appropriate training identified.

Year 3 – Sustained under-performance would trigger an extensive review of milestones identified in the development program. These may be more process oriented and would also evaluate the training and development of school staff specified earlier in this process. The school would also be asked to work more closely with evaluators or designated school improvement partners to improve performance.

Year 4 – A fourth year of under-performance would initiate a school renewal process, following successful programs implemented in education systems in other OECD countries (Araújo 2009; Goldstein et al. 1998; Gray et al. 1999; Hopkins 2005; Stewart 2006).

14. The level at which a school is identified as under-performing according to their value-added score is further discussed in Section 8.

An extensive review of the reasons for the under-performance and why the development program has not been successful would be undertaken with recommendations for extensive changes that may include the replacement of key staff, hiring successful educators to lead different areas of the school, and implementation of a new school philosophy. The school renewal could also encompass symbolic changes to the school's infrastructure such as renovations or new building works. But the emphasis would remain on improving the performance of the school, especially in the areas designated as under-performing.

Similar steps could be used by school boards in the Catholic and independent school sectors. This may be particularly important for parents if schools charging high fees are found to be poorly performing.

Regardless of the processes used in each school sector, to enhance accountability across the education system, particularly political accountability, COAG should track the success of each system (Government, Catholic and independent sectors in each state and territory) in turning around under-performing schools. This would complement current inter-state comparisons¹⁵ and hold jurisdictions accountable for turning around under-performing schools. Variation in the rate of success in turning around under-performing schools would highlight the impact of different practices. Successful practices would be identified and could be shared between sectors.

6.3. School autonomy

Holding school principals accountable for the performance of their school (and having that performance made public) requires they be empowered with sufficient autonomy to act on value-added measures. This is particularly important for school principals who as leaders and the public face of schools must answer to publications of school performance measures.

School principals should be granted autonomy to take decisions that lift student progress.¹⁶ Individual teachers have continually been shown to have the greatest impact upon student performance and school principals should be empowered to determine who teaches in their school (Caldwell & Harris 2008; OECD 2005b; 2009b).

Holding school principals accountable for the performance of their school (and having that performance made public) requires they be empowered with sufficient autonomy to act.

15. For example, see COAG Reform Council (2009) National Education Agreement: Baseline performance report for 2008, COAG Reform Council, Sydney.

16. This issue will be further explored in a forthcoming Grattan Institute education report.

This requires an important policy change in virtually every Australian state and territory. Increased school autonomy has been shown to be associated with increased school performance in Australia and internationally (Caldwell 2002; OECD 2007; Woessmann et al. 2009). Change is required to empower schools and give them the decision-making responsibilities to deliver improved performance. Australian school principals have already highlighted the need for these changes and made the calls for increased autonomy (Ferrari 2009a).

There is currently varied school autonomy within Australian states and territories (Caldwell & Harris 2008). Victoria has led Australia in devolving management to schools since the “Schools of the Future” policy initiative introduced in the mid 1990s: the vast majority of Victoria’s education funding is now subject to local management decisions. At the opposite extreme, schools within New South Wales have very little autonomy: budgeting and staffing decisions are taken centrally by the State’s Department of Education and Training (DEEWR 2007).

Policy recommendation:

Value-added measures of school performance should become an important benchmark in school evaluation. School evaluators should make their qualitative judgements of best practice based on accurate evidence. In addition:

- Developmental steps should be taken with under-performing schools (as identified by their value-added scores)
- COAG should document the success of each system (Government, Catholic and independent sectors in each state and territory) in turning around under-performing schools.

Policy recommendation:

School principals must be granted autonomy to take decisions that determine the performance of the school for which they are held accountable. Teachers have continually been shown to have the greatest impact upon student performance and school principals should be empowered to determine who teaches in their school.

7. Promoting school choice

Informed school choice is greatly advanced with the publication of measures of school performance (Gorard et al. 2001). Promoting choice within the school sector is an explicit policy objective as agreed by education Ministers at the MCEETYA Meeting in June 2009 (MCEETYA 2009). If parents and families have access to information on school performance then more informed decisions can be made in choosing the school that best suits each student's needs. In addition, families may become more actively engaged in schools if they have more information about school performance, either to help address poor performance or to become part of a high-performing school. This may be particularly important given the rising costs of attending some independent schools (Perkins 2010).

In Australia, school funding follows students. If high-performing schools attract more students, they receive more funds. If the benefits from promoting school choice and subsequent resource allocations are to have full effect, then the measures of school performance used by parents and families must be accurate. If not, then parents are not choosing the most effective schools for their children, and the resources that flow from these decisions are being misallocated. This increases the need for accurate measures of school performance and, as argued here, a move to value-added measures of student progress.

Given the Government's focus on school choice (MCEETYA 2009), it is important to understand that in Australia, as in many OECD countries, there are a number of factors hindering school choice and the full benefits of effective choice will not be realised as long as these obstacles remain (OECD 2006).¹⁷ Families cannot choose to send their child to any school. The main factors hindering school choice are financial, with many families unable to afford expensive independent schools or relocate to catchment areas for their favoured government schools. A number of schools reject students, particularly if they live outside of their local area, as they are already full and do not have the resources to accommodate additional students (OECD 2008).

If the benefits of more informed school choice are to be realised, then changes have to be made to increase the ability of parents to choose their favoured school (Hoxby 2003). The current focus on more informed school choice seems to ignore this issue. Effective school choice needs to exist for families before we worry about informing school choice.

17. This issue will be further explored in a forthcoming Grattan Institute education report.

8. Implementation issues

The discussion presented here provides a brief description of some of the key implementation issues from the recommendations presented in this report. This report has highlighted the benefits of value-added modelling but success depends upon effective implementation. There are numerous examples where poor implementation of systems of value-added modelling have resulted in poor outcomes (OECD 2008). This is not a full discussion of each step required in a complete implementation but builds on experiences, both positive and negative, gleaned from implementation in other education systems.¹⁸

Most implementation issues should be addressed during a pilot phase which would focus on choosing the appropriate value-added model. Other key issues discussed in this section include: the outcome measure upon which overall school performance is assessed; the presentation of school performance information; the need to greatly improve the quality and breadth of student background data to obtain accurate measures of school performance; and, a brief discussion of cost issues. This section does not explicitly detail the need for a comprehensive strategy for effective stakeholder engagement as it is considered that this is a theme running throughout this report.

School principals and teachers should be trained and empowered to use value-added performance information for school and teacher improvement and therefore be engaged at each stage of the development process.

8.1. Choice of value-added model

The introduction of value-added measures of school performance inevitably leads to questions of which model should be used. This question should be answered during the pilot phase which assesses the merits of each model run against existing data (or a subset thereof). The most appropriate model can be chosen which best suits the statistical and methodological criteria employed. This may focus on the variance produced with each model and the number of schools found to be statistically different from the average; the potential bias in the model and the handling of any missing data; and the reliability of estimates, particularly those for smaller schools.

It is important to recognise that model choice will not normally produce substantially different results. Analysis has shown that results are largely reproducible with different models (OECD 2008). This is not to say that model choice has no impact, merely that its impact should not be overstated. Steps should also be taken to ensure that small changes in results between models do not have negative impacts.

Estimates of schools labelled as under-performing should be the subject of further analysis. Analysis of outliers in the data, missing data, and the possibilities of misspecification should be analysed. This may include alternate models being run for schools categorised as under-performing to ensure that spurious results do not trigger misguided actions.

18. For a more complete discussion of implementation issues see OECD (2008), *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*, OECD, Paris.

8.2. Choosing a single measure of school performance

It is important to identify early in the process the value-added measures upon which school performance will be based. NAPLAN assesses literacy and numeracy across five domains of learning at 4 year levels of school education: a combined primary-secondary school would therefore have at least 15 value-added scores. It is necessary to identify the measure that would trigger actions to improve school performance. A single measure provides consistency across years and may also improve policy decisions such as school funding models targeting under-performing schools.

Choosing a single measure of school performance requires that policy-makers recognise the impact this will have on schools and teachers. It is also important to recognise that a value judgement is being made. First, measuring performance in specific subjects, students or standards encourages school principals and teachers to focus on these areas, especially if there are extensive developments and accountability actions stemming from the value-added scores. Second, it requires a value judgement about where the education system should focus its resources. For example, focusing on low performing students over other students requires a value judgement that governments may wish to make if they believe the social and economic benefits are greater. It is argued here that all value-added scores should be calculated and made available to schools and the general public. It is beneficial for more information to be made available to schools for their internal development and to feed other policy objectives.

Nevertheless, it is important that a single measure be identified as the prime measure of school performance. There must be clarity about specific actions, especially in classifying a school as under-performing and undertaking actions to address that under-performance. This should be the overall value-added measure for a school that is the average of value-added for all subject areas assessed across year levels. Focusing on value-added across all assessments avoids the potentially negative consequences of focusing resources on specific areas of school education.

8.3. A 3-year moving average of school performance

This report has argued that value-added measures of school performance should be used as a basis for actions that benefit school improvement. Actions such as schools being placed 'on notice' and school renewal programs have also been discussed. Given the impacts of these initiatives it is essential that the measures of school performance be reliable. Research has shown that there can be year-to-year volatility in school's value-added results (Jakubowski 2007; Ray 2007; van de Grift & Houtveen 2006) and that this volatility is often greater for small schools (OECD 2008).¹⁹

Volatility in school performance between single years is unlikely to reflect large changes in performance and may be partly due to random noise in the data and/or the estimation process (Kane & Staiger 2002). Given these findings and the need to produce accurate measures of school performance it is recommended that schools value-added scores be calculated as a 3-year moving average. This smooths year- on-year fluctuations to create a more stable and reliable measure of school performance.

The instability of results of small schools means that value-added results calculated for schools where the sample of students is less than a specified number (determined through testing in the pilot phase) should not hold implications for teachers and schools. For example, small schools should not be used for actions such as school renewals that have implications for teachers and school principals. Instead, they should be used as additional information that complements other evaluative information, with caveats to inform readers of potential problems with the data.

As with other statistical and methodological issues, substantial testing in the pilot phase of the implementation process will reveal the extent of volatility of results with different models. The appropriate sample size for the classification of small schools can then best be ascertained.

8.4. Increasing accuracy and effectiveness with annual NAPLAN testing

We advocate for the eventual introduction of annual NAPLAN assessments for students between year 3 and year 10. The result of this would be value-added measures available for each student in most years of their school education. This would increase the opportunities for schools to identify the cause of students progressing at faster and slower rates, leading to more timely improvements in instruction and school programs. Importantly, more frequent student assessments permits more accurate value-added measures of school performance (Raudenbush 2004; Sanders et al. 1997).

Currently, school performance data is limited in enabling change within schools. As students are only tested every second year, it is more difficult to identify the true causes of student progress. For example, analysing school value-added between years 7 and 9 does not tell us if progress was greater in year 7 or year 8. Greater precision is required to facilitate school improvement initiatives. Current efforts to improve instruction are negated as value-added can only be calculated once in students' primary education (between years 3 and 5) and twice in students' secondary education (between years 7 and 9, and year 9 and the final year of secondary school).

19. It should be noted that if volatility between school value-added scores is a large concern then this should be incorporated into decisions concerning model choice in the implementation stage. For example, a feature of random effects models such as that used in England are that the school performance residuals are 'shrunk' to the average, particularly for smaller schools (Ray, 2007). This reduces volatility but can increase bias in the results which should also be considered when comparing the efficacy of different value-added models.

8.5. Presenting value-added measures of school performance

Section 5.2 provides examples of how value-added information can effectively be used in schools to inform decision-making and improve instruction. Presentation of value-added information to the public can complement the existing dissemination of school reporting information through the *'My School'* website. Value-added measures should be presented for each NAPLAN learning domain for each year level assessed. A value-added score for students in their final year of secondary school would add to this picture and would be particularly important for students (as performance in these years largely determines university entrance). These would be complemented by an overall school value-added score that is the average of these scores.

Value-added scores complement and should be presented alongside NAPLAN scores to provide as much information as possible to inform all stakeholders. However, value-added measures should be emphasised as the true measure of performance. This would severely dilute the possibilities of league tables being created based on raw test scores. This is more likely with the current system that does not produce an alternate performance measure.

An additional benefit of presenting schools' NAPLAN scores alongside value-added scores is that it would diminish the prospect of schools 'gaming' the system that has been a problem in some education systems (Jacob 2002). Value-added measures of school performance emphasise student progress between NAPLAN assessments and, as recommended here, students' grades in their final year of secondary school. As school performance is measured on the progress made between these assessments, there is an incentive for schools to score poorly on an initial assessment. For example, a secondary school could discourage their students from performing well on the year 7 NAPLAN assessments to artificially accentuate the progress made to the year 9 assessment and then to the final year secondary school. By publishing NAPLAN scores as well as value-added scores, schools 'gaming' the system would have published NAPLAN scores that are artificially low. It is safe to assume that no school wants to have lower NAPLAN scores or be known as having low-performing students.

An important aspect that needs to be considered is the scale of the value-added score. As discussed earlier in this report, value-added is a relative measure with a positive score indicating a school has added more value than the average and a negative score indicating that the school has contributed less than the average to student progress. Difficulties can exist in presenting negative scores as it can give the mistaken impression that students in the school are going backwards. Students in a school with a negative value-added score are still progressing but the contribution of the school to their rate of progress is less than the expected average. Given the possibility of confusion stemming from the relative nature of value-added scores, it is recommended that school value-added scores are converted to a scale with a mean of 100. Therefore, a school with a zero value-added score is converted in the new scale to a value-added score of 100. Schools would therefore no longer have negative scores. They would simply score less than 100. Similar changes were made in England to reduce confusion over this matter (Ray 2007).

8.6. The need for accurate student-level data

An integral component in developing accurate measures of school value-added is that the data used in the modelling is reliable, accurate and has the required coverage to measure school performance. Clearly, the quality of the student assessments is of utmost importance and this must be maintained over time to ensure that we can measure student progress (Koretz 2005). Given their high quality, the NAPLAN assessments are not questioned here. However, the reliability of student-level data is also very important to the development of accurate measure of school performance: in Australia there are substantial problems with the quality and coverage of this data that needs to be addressed.

A lack of student background contextual data restricts both the accuracy and utility of the current measures of the performance of Australian schools. Substantial evidence in numerous countries has shown that the accuracy of school performance measures relies upon the quality of the student background data included in the model (Goldstein et al. 2008).

This encompasses not only the extent that socio-economic background data is included in the modelling (Ladd & Walsh 2002) but also which specific characteristics best capture the factors affecting student performance that are outside the control of the school (Ferrão 2007; Hægeland & Kirkebøen 2008).

The school performance measures calculated by the Federal Government rely on census data of where a student resides to control for differences in students background that would influence their performance in the NAPLAN assessments. This is a crude measure of student background that does not properly account for heterogeneity within and between schools. It does not account for the array of socio-economic background characteristics that can influence student performance. However, they are the best that is currently available and still provide valuable output. Naturally, they are better than using no background measures. The data used by the Victorian Government to create their intake-adjusted school performance measures are better than that used by the Federal Government, not least because it utilises student and school-level data instead of census data. However, more detail is required on student socio-economic status to better control for differences in student backgrounds to accurately measure school performance.

An important additional benefit from improved student background data is that the greater the background data available, the more that educators can use the data to help improve the education offered to specific groups of students (e.g. immigrant or low socio-economic backgrounds). As an example, if we consider a student taking the NAPLAN Year 5 numeracy assessment, there are numerous factors that will influence that student's performance in that assessment.

These factors can roughly be grouped into individual, family, neighbourhood and community, and school factors that influence student performance on this assessment.

Individual factors include characteristics such as gender, mobility between schools and of course innate intelligence that influence student performance (OECD 2007; Ray 2008; Thomson & De Bortoli 2008). While there is not normally great variation in the age of students within year levels, such variation can significantly affect performance (Hægeland et al. 2005; Ray 2006). A host of familial characteristics ranging from family size and marital status to education and employment influence student performance (Haveman & Wolfe 1995; Lissitz et al. 2006). Given the limitations of collecting data and the need to keep the modelling relatively simple, value-added analyses in OECD countries have usually focused on parents' education, occupation and employment status but can also include factors such as divorce that can influence students' academic progress (OECD 2008). Neighbourhood and community factors are often more difficult to measure but focus on the characteristics of people in students' neighbourhood and peer group (Ginther et al. 2000).

Measures of these factors, or proxies of them, need to be included in a statistical estimation of school performance as we have to 'control' for these factors to isolate the effect of a school on student performance. This is not a simple process given the array of factors that influence student performance and the difficulties in identifying them, let alone measuring them. But if not properly addressed, problems of bias and misspecification are often created (OECD 2008).

So for each student, a statistical estimation of their performance may include their: age; gender; indigenous status; whether or not they are new to a school; family education and occupation; country of birth and migration background; language preferences; grade repetition; and student learning difficulties. These factors affect the performance of students but are outside the control of the school. They have been critical in systems in other OECD countries to accurately measure school performance (Ferrão 2008; Hægeland & Kirkebøen 2008).

Student-level information on these areas should be collected to create more accurate measures of school performance and target programs for minority groups. Such data collections have been successfully implemented in a number of countries without placing an excessive burden on schools or infringing on people's privacy. Similar data has been successfully collected in Victoria and this could be easily extended (DEECD 2009).

If we consider the above list of student background factors that should be included in a contextualised value-added measure of school performance, it is clear that:

- The crude measures used in the ‘*My School*’ website to create like-school groups are inadequate to accurately isolate and therefore measure school performance;
- The required student background contextual information can be readily collected at the student-level and has been shown in a number of education systems not to place an onerous burden on schools; and
- Measures of the value-added of schools to the progress of students of different backgrounds (as measured with this student background data) would greatly assist efforts to develop and monitor policies, both at the system and school level, to improve the performance of students from different backgrounds (e.g. analysing the impact of grade repetition on the progress of students with a non-English language speaking background).

If student background characteristics are not properly controlled for in the statistical estimation, then the measures of school performance become biased. This may inaccurately provide a school performance measure that overstates or under values the true school performance. The magnitude of this problem is greatly reduced with value-added modelling but it is clear that a student background questionnaire should be completed for each Australian student that provides sufficient information to allow for accurate measures of school performance. Opponents of such a move need to recognise that without this information we are harming and potentially punishing schools serving poorer communities. Moreover, we are preventing educators from improving school education for minority and disadvantaged groups. Arguments of placing an undue burden on schools simply do not hold given the costs of such a questionnaire compared to the advantages of improving the system.

8.7. Cost

A final issue that must be considered in every policy initiative or set of recommendations is cost. The costs and benefits of all education policies and programs should be continually monitored.

While detailed cost analysis for each state and territory is outside the scope of this report, some issues should be considered such as education and training expenditures and costs associated with the development and dissemination of value-added measures.

The costs in the move to value-added modelling are relatively low. The main costs come from educating and training for school principals and teachers to incorporate value-added measures into their decision-making and instruction. Numerous examples are available of school principals and teachers being trained in value-added modelling in easily reproducible formats (OECD 2008). Jakubowski (2008) illustrates how short training sessions were effectively used to train Polish teachers to not only use the data but calculate their own value-added scores with statistical modelling packages. Such extensive training may not be necessary but it illustrates the point that this is not a confusing statistical concept with a level of complexity that prevents effective dissemination and use. Such training incurs a cost that is not insignificant but can be reduced if existing expenditures on education and training are targeted in these areas.

This may be particularly useful considering the benefits to students achievement discussed in Section 5 (Black & William 1998; EPPI 2002; Griffin et al. *forthcoming*; OECD 2005a; Taylor et al. 2005).

Costs associated with the modelling itself and the implementation of an effective user-friendly information technology system have not been large in other education systems. For example, in England, a small team of 3-5 people run the value-added modelling for all schools in England. This team also assumes a number of additional responsibilities concerning the dissemination of performance results, the use of RAISE Online (the information technology system used by schools to analyse performance information in England), and the key stage assessments of students (OECD 2008).

It should also be remembered that the greatest costs in the measurement of school performance have already been committed in Australia.

The development and implementation of the NAPLAN assessments is the largest cost component in this process by a substantial margin. The recommendation to extend to annual testing of students between years 3 and 10 would add to these costs. However, as the infrastructure for developing and holding NAPLAN tests is already in place, the cost implications are reduced. The benefits would be considerable and given the investments already made in the NAPLAN assessments and the objectives to raise school performance, the cost-benefit analysis of the recommendations is likely to be very positive.

9. Conclusion

Publishing schools' NAPLAN scores on the 'My School' website is a large step forward. However, for the objective of measuring school performance, school value-added scores should replace the current system that relies on like-school groups, that is prone to inaccuracies and may be biased against schools serving lower socio-economic communities. Inaccurate measures of school performance will not have the intended benefits of increasing school accountability and school choice and negate the substantial opportunities for improving school effectiveness.

Value-added analysis allows policy makers and educators to track student progress through their schooling, rather than relying on simplistic snapshots of performance at a single point in time. Regardless of students' social, economic or language backgrounds, or the school they attend, the focus should be on the progress, or lack thereof, made by all students.

Measures of value-added school performance should not be viewed as an end in themselves. Rather, they are a basis for action. Value-added measures will form an effective evidence base for informed policy development, and improve instructional practices and school effectiveness.

This report advocates for the adoption of value added measures of school performance throughout Australia. Value-added measures of school performance provide greater accuracy and fairness to schools serving different student populations. Further recommendations are also made that will vary in their impact between education systems in Australia that include:

- School principals and teachers should be empowered to use value-added measures to improve instruction and school programs. To achieve this:
 - A user-friendly information technology system should be developed that allows school principals and teachers to better analyse and then act upon their own performance data;
 - Education and training to incorporate performance assessment into instruction and school programs should be provided;
 - Resources should be made available for teachers and schools to develop programs based on value-added measures and disseminate best practice.
- Value-added measures of school performance should become an important benchmark in school evaluation. School evaluators should make their qualitative judgements of good practice in the context of value-added performance measures;
- Value-added measures of student progress should be the basis for categorising schools as under-performing. Developmental steps should be explicit, with additional support for under-performing schools; and
- School principals must be granted autonomy to effectively lead the school for which they are being held accountable. Individual teachers have continually been shown to have the greatest impact upon student performance and school principals should be empowered to determine who teaches in their school.

10. References

Araújo, M. 2009, 'A Fresh Start for a 'failing school'? A qualitative study', *British Educational Research Journal*, vol. 35, no. 4, pp. 599-617.

Australian Education Union 2009, *Parents, Teachers and Principals Call for Action on School League Tables*, media release, Australian Education Union, Victoria, November 2009.
From <http://www.aeufederal.org.au/Media/MediaReleases/2009/1711.pdf>.

Ballou, D., Sanders, W & Wright, P 2004, 'Controlling for Student Background in Value-Added Assessment of Teachers.' *Journal of Educational and Behavioural Statistics*, vol. 29.

Bethell, G. 2005, *Value-Added Indicators of School Performance: The English Experience Anglia Assessment*. Battisford, Suffolk, England: Unpublished report.

Black, P. & William, D 1998, 'Assessment and Classroom Learning', *Assessment in Education: Principles, Policy and Practice*, CARFAX, Oxfordshire, vol.5, No. 1, pp. 7-74.

Braun, H.I. 2005, *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Policy Information Perspective. ETS.

Braun, H. I., Qu Y. & Trapani. C. S. 2008. *Robustness of Value-added Analysis of School Effectiveness*. ETS RR-08-22. Princeton, NJ: Educational Testing Service.

Caldwell, B. 2002, 'Autonomy and Self-management: Concepts and Evidence.' In T. Bush and L. Bell, *The Principles and Practice of Educational Management* (pp. 34-48). London: Paul Chapman.

Caldwell, B J. & Harris J L 2008, *Breakthrough in governance [electronic resource]* / Brian Caldwell and Jessica Harris Educational Transformations, Brighton, Vic. Sourced from <http://www.ssat-inet.net>.

Choi, K., P. Goldschmidt & K.Yamashiro. 2005, 'Exploring Models of School Performance: From Theory to Practice.' In J. H. (eds.), *Yearbook for the National Society for the Study of Education*, Malden, Massachusetts: Blackwell, vol. 104, no. 2.

COAG Reform Council 2009, *National Education Agreement: Baseline performance report for 2008*, COAG Reform Council, Sydney.

DEECD 2009, *Measuring Performance Fairly: Developing intake adjusted performance measures in Victorian government schools*, Published by the Data, Outcomes and Evaluation Division Office for Children and Portfolio Coordination, Department of Education and Early Childhood Development, Victorian Government, Melbourne, November 2009.

Department of Education, Employment and Workplace Relations 2007, *Principal Autonomy Research Project*, Australian Government, 5 December 2007.

Department of Education, Employment and Workplace Relations 2009a, *Using ICSEA, My School Factsheet*, www.myschool.edu.au, Australian Government, December 2009.

Department of Education, Employment and Workplace Relations 2009b, *About My School*, My School Factsheet, www.myschool.edu.au, Australian Government, December 2009.

Department of Education, Employment and Workplace Relations 2009c, *About NAPLAN*, My School Factsheet, www.myschool.edu.au, Australian Government, December 2009.

Dixit, A. 2002, 'Incentives and Organisations in the Public Sector: An Interpretive Review'. *Journal of Human Resources*, vol. 37, no. 4, pp. 696-727.

Donnelly, K. 2010, 'I Was Wrong on League Tables for Schools', *The Age*, January 15 2010, can be accessed at: <http://www.theage.com.au/opinion/contributors/i-was-wrong-on-league-tables-for-schools-20100115-mbq5.html>.

Dudley, P. 1999, 'Using Data to Drive Up Standards: Statistics or Psychology?' In C. Conner (ed.), *Assessment in Action in the Primary School*. London: Falmer Press.

Doran, H. C. and L. T. Izumi. 2004, *Putting Education to the Test: A Value-Added Model for California*. San Francisco: Pacific Research Institute.

EPPI 2002, *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*, Institute of Education, University of London, June.

Ferrari, J. 2009a, 'Principals Seek to Hire and Fire', *The Australian*, 16 December 2009, p.1.

Ferrari, J. 2009b, Backdown on National School 'league tables', *The Australian*, 13 May 2009, can be accessed at <http://www.theaustralian.com.au/news/nation/backdown-on-school-league-tables/story-e6frg6nf-1111116321624>.

Ferrão, M. 2007, *Sensitivity of VAM Specifications: Measuring Socio-Economic Status: A Background Report for the OECD Project on the Development of Value-added Models in Education Systems*. Warsaw.

Fitz-Gibbon, C. 1997, *The Value Added National Project Final Report: Feasibility Studies for a National System of Value-Added Indicators*. London: School Curriculum and Assessment Authority.

Gillard, J 2009a, *After the Crisis: Hard Reform for a Better Future*, Speech to *The Australian/Melbourne Institute Conference*, 5 November 2009, Melbourne.

Gillard, J. 2009b, Speech to the National Public Education Forum, Canberra, 27 March 2009., can be accessed at http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article_090327_173128.aspx.

Ginther, D., Haveman R. & Wolfe B. 2000, 'Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are the Relationships?' *Journal of Human Resources*, vol. 34, no. 4, pp. 603-42.

Goldstein, J., Kelemen, M., & Koski, W. 1998, "*Reconstitution in Theory and Practice: The Experience of San Francisco*", paper presented at the Annual Meeting of the American Educational Research Association, San Diego, April 13-17, 1998.

Goldstein, H., Kounali D. & A. Robinson A. 2008, 'Modelling Measurement Errors and Category Miss-classifications in Multilevel Models.' *Statistical Modelling*, vol. 8, no. 3, pp. 243-261.

Gorard, S., Fitz J., & Taylor C., 2001, 'School Choice Impacts: What Do We Know?' *Educational Researcher*, vol. 30 no. 7, pp.18-23.

Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S. & Jesson, D. 1999, *Improving Schools: Performance and Potential*, Open Universities Press, Buckingham.

Griffin, P. 1990, 'Profiling literacy development: Monitoring the accumulation of reading skills.' *Australian Journal of Education*, vol. 34, no. 3, pp. 290-311.

Griffin, P., Murray, L., Care, E. & Thomas, A., & Perri P. *forthcoming* 'Developmental Assessment: Lifting Literacy through Professional Learning Teams', *Journal of Assessment in Education*.

Hægeland, T., Kirkebøen, L., Raaum O., & Salvenes K., 2005, *School performance indicators for Oslo, Reports 2005/36*. Statistics Norway.

Hægeland, T. & Kirkebøen K., 2008, *School Performance and Value-Added Indicators – What is the Importance of Controlling for Socioeconomic Background?* A Background Report for the OECD Project on the Development of Value-added Models in Education Systems.

Hallinger, P. & Heck R., 1998., 'Exploring the Principal's Contribution to School Effectiveness: 1980-1995', *School Effectiveness and School Improvement*, vol. 9 no. 2, pp.157-191.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. 2005., *The new instructional leadership: Creating data-driven instructional systems in schools*. (WCER Working Paper No. 2005-9). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Haveman, R. & B.Wolfe. 1995., The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, vol. 33, pp. 1829-1878.

Hanushek, E.A. & Woessmann L. 2007, *The Role of Education Quality in Economic Growth*, World Bank Policy Research Working Paper 4122, February 2007.

Hopkins, D. (ed) 2005, *The Practice and Theory of School Improvement: International Handbook of Educational Change*, Springer, Dordrecht.

Hoxby, C. 2003, *The Economics of School Choice*, *National Bureau of Economic Research Conference Report*. University of Chicago Press, Chicago.

Jacob, B. 2002, *Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools*. NBER Working Paper No. 8968. Cambridge, MA.

Jakubowski, M. 2008, Implementing Value-Added Models of School Assessment. *RSCAS Working Papers 2008/06*, *European University Institute*.

Jakubowski, M. 2007. Volatility of Value-Added Estimates of School Effectiveness: A Comparative Study of Poland and Slovenia. *Paper presented to the Robert Shurman Centre for Advanced Studies, European University*. Florence.

Kane, T.J. & Staiger D. O. 2002, 'Volatility in School Test Scores: Implications for Test-Based Accountability Systems' In D. R. (Ed.), *Brookings Papers on Education Policy* (pp. 235-269). Washington, DC: Brookings Institution.

Klieme, E., Lipowsky F., Rakoczy K., & Ratzka N. 2006, "Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht: Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts 'Pythagoras'", in: M. Prenzel & L. Allolio-Näcke (eds.), *Untersuchungen zur Bildungsqualität von Schule, Abschlussbericht des DFG-Schwerpunktprogramms*, Waxmann, Munster pp. 128-146.

Koretz, D. 2005. Alignment, High Stakes, and the Inflation of Test Scores. In J. L. Herman & E. H. Haertel (ed.), *Uses and Misuses of Data for Educational Accountability and Improvement* Malden, MA: NSSE, (pp. 99-118).

Ladd, H.F. 2007, 'Holding Schools Accountable Revisited,' 2007 Spencer Foundation lecture in Education Policy and Management, Association for Public Policy Analysis and Management.

Ladd, H. and Figlio, D. 2008, 'School Accountability and Student Achievement', in Helen F. Ladd & Edward B. Fiske, (eds.) *Handbook of Research on Education Finance and Policy*, Routledge.

Ladd, H. F. and Walsh R. P. 2002, 'Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right.' *Economics of Education Review*, vol. 21, pp. 1-17.

Leigh, A. 2009, 'Estimating teacher effectiveness from two-year changes in students' test scores' *Economics of Education Review*, doi: 10.1016/j.econedurev.2009.10.010.

Lissitz, R., Doran H., Schafer W. & Willhoft J. 2006, 'Growth Modelling, Value-Added Modelling and Linking: An Introduction'. In R. Lissitz, *Longitudinal and Value-Added Models of Student Performance*, Mapple Grove, MN, JAM Press, pp. 1-46.

Masters, G., Rowley, G., Ainley, J. & S.T. Khoo 2008, 'Reporting and Comparing School Performances' Paper prepared for the MCEETYA Expert Working Group, December 2008.

McCaffrey, D. F., Lockwood, J. R., Mariano, L. T. & Setodji C. 2005, 'Challenges for value-added assessment of teacher effects.' In R. Lissitz (Ed.) *Value added models in education: Theory and practice*. Maple Grove, MN: JAM Press.

McCaffrey, D. F., Lockwood J. R., Koretz D. M. & Hamilton L. S. 2003, *Evaluating Value-Added Models for Teacher Accountability*. The RAND Corporation, Santa Monica, CA.

McCaffrey, D. M., Lockwood J. R., Koretz, D., Louis T. A. & Hamilton L., 2004, 'Models for Value-Added Modelling of Teacher Effects.' *Journal of Educational and Behavioral Statistics*, vol. 29 no. 1, pp. 67-101.

Ministerial Council on Education, Employment, Training and Youth Affairs 2008 *Melbourne Declaration on Educational goals for Young Australians*, Ministerial Council on Education, Employment, Training and Youth Affairs, Melbourne, December, 2008.

Ministerial Council on Education, Employment, Training and Youth Affairs 2009a, *National Assessment Program – Literacy and Numeracy: NAPLAN Summary Report*, Ministerial Council on Education, Employment, Training and Youth Affairs, Canberra, can be accessed: http://www.naplan.edu.au/verve/_resources/2009_NAPLAN_Summary_Report.pdf.

Ministerial Council on Education, Employment, Training and Youth Affairs 2009b, *Principles and Protocols for Reporting Schooling in Australia*. Can be accessed at: http://www.mceecdya.edu.au/verve/_resources/Principles_and_Protocols_2009.pdf.

Mokhtari, K., Rosemary, C.A., & Edwards, P.A. 2007, 'Making instructional decisions based on data: What, how, and why.' *The Reading Teacher*, vol. 61 no. 4, pp. 354–359.

National Research Council. (2010). *Getting Value Out of Value-Added: Report of a Workshop*. Committee on Value-Added Methodology for Instruction Improvement, Program Evaluation, and Accountability. H.I. Braun, N. Chudowsky, and J. Koenig, (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC : The National Academies Press.

NASBE 2005, *Evaluating Value-Added: Findings and Recommendations from the NASBE Study Group on Value-Added Assessments*. National Association of State Boards of Education, Alexandria, VA, USA.

OECD 2005a, *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.

OECD 2005b, *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OECD, Paris.

OECD 2006, *Demand-Sensitive Schooling? Evidence and Issues*, OECD, Paris.

OECD 2007, *PISA 2006: Science Competencies For Tomorrow's World*, OECD, Paris.

OECD 2008, *Measuring Improvements in Learning Outcomes: Best practices to Assess the Value-Added of Schools* OECD, Paris.

OECD 2009a, *Education at a Glance - OECD Indicators 2008*, OECD, Paris.

OECD 2009b, *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD, Paris.

O'Day, J. 2002, Complexity, Accountability, and School Improvement. *Harvard Educational Review*, vol. 72, no. 3, pp. 293-329.

Patty, A. & Gilmore H. 2009, 'School still top of class', *Sydney Morning Herald*, 17 December 2009, p 1.

Perkins, M. 2010, 'Private School Fees race Ahead of Inflation,' *The Age*, 7 January 2010, accessed at: <http://www.theage.com.au/national/education/private-school-fees-race-ahead-of-inflation-20100106-lubx.html>.

Perkins, M., 2009, 'Full marks for Mount Scopus in VCE results', *The Age*, 17 December 2009, p9.

Phillips, G., McNaughton S., & MacDonald, S. 2004, 'Managing the mismatch: Enhancing early literacy progress for children with diverse language and cultural identities in mainstream urban schools in New Zealand.' *Journal of Educational Psychology*, vol 96, no. 2, pp. 309-323.

Ponisziak, P. M. & Bryk A. S. 2005, 'Value-Added Analysis of the Chicago Public Schools: An Application of Hierarchical Models'. In R. L. (Ed.), *Value Added Models in Education: Theory and Applications*. Maple Grove, MN: JAM Press.

Raudenbush, S. and Willms J. D. 1995, 'The Estimation of School Effects.' *Journal of Educational and Behavioral Statistics*, vol. 20, pp. 307-335.

Raudenbush, S. W. 2004, *Schooling, Statistics, and Poverty: Can We Measure School Improvement?* Educational Testing Service, Princeton, NJ, USA.

Ray, A. 2006, *School Value Added Measures in England: A Background Report for the OECD Project on the Development of Value-Added Models in Education Systems*, www.dcsf.gov.uk/rsgateway/DB/RRP/u015013/index.shtml.

Ray, A. 2007, *The Volatility of Value-Added Scores: A Background Report for the OECD Project on the Development of Value-Added Models in Education Systems*, unpublished, OECD, Paris.

Ryska, R. 2006, *Value-added Modelling in the Czech Republic: A Background Report for the OECD Project on the Development of Value-added Models in Education Systems*, unpublished, OECD, Paris.

Sanders, W., Saxton A., & Horn B. 1997, 'The Tennessee Value-Added Assessment System: A Quantitative Outcomes-Based Approach to Educational Assessment.' In J. M. (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluational Measure?* Corwin Press, Inc, Thousand Oaks, CA, pp.137-162.

Stewart, B. (2006) *Value-Added Modelling: The Challenge of Measuring Educational Outcomes*, Carnegie Institute.

Sydney Morning Herald 2009, 'The Top 100: Schools ranked by percentage of students who scored more than 90 in their HSC Courses', *Sydney Morning Herald* 17 December 2009, p 2.

Taylor, B. M., Pearson, P. D., Peterson, D., & Rodriguez, M. C. 2005, 'The CIERA school change framework: An evidence based approach to professional development and school reading improvement.' *Reading Research Quarterly*, vol. 40, no. 1, pp. 40-69.

Tekwe, C., Carter R., Ma C., Algina J., Lucas M. & Roth J. 2004, 'An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance.' *Journal of Educational and Behavioral Statistics*, vo. 29 no. 1, pp. 11-36.

The Age 2009, '2009 Your Schools VCE Performance', Lift out, *The Age*, 17 December 2009.

The Hobart Mercury (2009) 'Tassie High School Rankings', *The Hobart Mercury* 6 May 2009 http://www.themercury.com.au/article/2009/05/06/71261_tasmania-news.html, Retrieved 13 January 2010.

Timperley, H. S., & Robinson, V. J. M. 2001, 'Achieving school improvement through challenging and changing teachers' schema.' *Journal of Educational Change*, vol. 2, pp. 281-300.

Tomazin, F. 2009, 'League Table Slammed as a Disaster', *The Age*, 28 May 2009 (can be accessed at www.theage.com.au/national/education/league-tables-slammed-as-a-disaster-20090527-bnni.html).

Thomson, S. & De Bortoli, L. 2008, *Exploring Scientific Literacy: How Australia measures up: the PISA 2006 survey of students' scientific, reading and mathematical literacy skills*, Australian Council for Educational Research, Camberwell, Vic, Aust.

van de Grift, W. 2007, *Reliability and Validity in Measuring the Added Value of Schools: A Background Report for the OECD Project on the Development of Value-Added Models in Education Systems*, OECD, Paris.

van de Grift, W. 2001, 'Waarom hebben sommige scholen onderprestatie?' [Why do some schools suffer from underachievement?] *Basis school management* vol 15 no. 1, pp. 1-9.

van de Grift, W.J.C.M. and Houtveen A.A.M. 2006 'Under-performance in Primary Schools', *School Effectiveness and School Improvement*, Vol. 17, No. 3, September 2006, pp. 255-273.

Vygotsky, L. S. (1986). *Thought and language*. MIT Press, Boston.

William, D., Lee, C., Harrison, C. & Black P. 2004, 'Teachers developing assessment for learning: impact on student achievement,' *Assessment in Education: Principles, Policy and Practice* (March 2004), pp. 49-65.

Willms, J., & Raudenbush, S. 1989, 'A longitudinal hierarchical linear model for estimating school effects and their stability.' *Journal of Educational Measurement*, vol. 26 no. 3 pp. 209-232.

Woessmann, L., Ludemann C., Schultz S., & West M. R., 2009, *School Accountability, Autonomy, and Choice Around the World*, Edward Elgar Publishing, IFO, Munich.