Reporting School Results and Student Achievement

Submission to the Senate Education Committee Inquiry into the Administration and Reporting of NAPLAN

Part 1

SAVE OUR SCHOOLS

June 2010

http://www.saveourschools.com.au

Table of Contents

Key Points

1. A review of the major research studies on reporting school results, school accountability measures and policies to promote greater choice and competition between schools shows that the Govenment's claim that publishing school results will improve student achievement is unsubstantiated. At best, the evidence is mixed. This is not a robust foundation for education policy.

2. Studies of the impact of reporting school results show show no significant effect on student achievement.

3. The evidence from the major studies of "high stakes" accountability measures (which include reporting school results, rewards and sanctions for schools, exams for students and grade promotion standards) is mixed. Some find positive effects on student achievement, others find no effect or negative results.

4. The most frequently cited studies showing a positive effect from "high stakes" accountability over-state the effect because of technical flaws in the measurement of the effect.

5. Much of the positive effect of school accountability measures shown in key studies is due to student-based requirements such as meeting grade promotion standards, passing end-of-course exams and graduation exams. The effect of these requirements on student achievement is larger than the effect of rewards and sanctions relating to school performance.

6. There is substantial evidence that the positive results in some studies may be due, at least in part, to schools manipulating their results in various ways such as re-classifying students as special education students so they can be exempted from high stakes tests, suspending low achieving students during the testing cycle and outright cheating.

7. The major academic reviews of "high stakes" accountability studies conclude that the evidence is mixed and provides little scientific foundation for these policies. This has been acknowledged by the chief executive of the Australian Curriculum, Assessment and Reporting Authority, Peter Hill, thus contradicting his Minister's claims.

8. The weight of evidence from the best designed and most comprehensive of studies on increasing choice and competition between schools is that it does not improve student achievement once student background characteristics are taken into account.

9. Former advocates of "high stakes" school accountability measures and greater choice and competition between schools such as Chester Finn and Diane Ravitch have been forced to admit that these policies have not worked.

Summary

The Prime Minister and the Federal Education Minister claim that reporting school NAPLAN results on the My School website will lead to improved school performance.

A review of the major research studies on reporting school results, school accountability measures and policies to promote greater choice and competition between schools shows that the Govenment's claim is unsubstantiated. At best, the evidence is mixed. This is not a robust foundation for education policy.

The weight of evidence of the few studies that have analysed school reporting separately from other school accountability measures is that it fails to increase student achievement. Indeed, one of the most frequently cited studies on school accountability concluded that reporting school results itself has no effect on student achievement [Hanushek & Raymond 2005].

The evidence from the major studies of so-called "high stakes" accountability measures (which include reporting school results, rewards and sanctions for schools, exams for students and grade promotion standards) is mixed. Some find positive effects on student achievement, others find no effect or negative results. Often the results differ within studies as well as between studies. Nearly all of this evidence comes from studies conducted in the United States over the past decade or so.

The effects on reading and mathematics achievement were sometimes modestly negative and sometimes modestly positive. The effects varied among studies depending on the nature, types, and timing of accountability policies. They varied also among subjects, grades, and time periods chosen for the analysis. The methodology used also affected the results.

There are several reasons to believe that the positive results shown in some studies are over-stated.

The two most frequently cited studies which show that accountability measures increase school achievement [Carnoy & Loeb 2002; Hanushek & Raymond 2005] significantly over-stated the effect. The studies measured the effect as a proportion of the variation in state results instead of the variation in student results, which is much larger and is the more relevant benchmark. When measured against the variation in student achievement the positive effect in these studies is shown to be very small.

Most of the studies showing a positive effect from high stakes accountability include student as well as school accountability measures. This also causes the effect to be over-stated.

The studies take into account school-based measures such as reporting school results, monetary rewards for improved performance and sanctions for low performance such replacing the principal or teachers, re-organising or closing schools and permitting students to enrol elsewhere. They also include requirements for students such as grade promotion standards and tests; end-of-course exams; and a high school graduation exam.

Student-based requirements like these have a significant effect on the final results of many high stakes accountability studies; and their effect is larger than the effects of rewards and

sanctions relating to school performance. Indeed, a study which re-worked the data used in one of the most frequently cited studies finding positive results from high stakes accountability found that the results were attributable to the student requirements.

There is also substantial evidence that the positive results in some studies may be due, at least in part, to schools manipulating their results in various ways. Studies show that schools respond to high stakes accountability tests by moving resources from subjects not tested to the subjects that are tested, re-classifying students as special education students so they can be exempted from high stakes tests, suspending low achieving students during the testing cycle and outright cheating.

A few studies of the impact of high stakes accountability regimes have examined their distributional consequences. They found that students just below proficiency standards tend to make greater test score gains as schools tend to concentrate resources on improving the results of this group. There is conflicting evidence concerning the effects on students scoring at either the highest or lowest end of the performance spectrum. There is also little evidence to suggest that accountability measures reduce achievement gaps between low and high income students and between white students and those from other racial groups.

Some studies using international test results have found small positive effects of school accountability measures on student achievement. However, these studies have significant methodological caveats attached to them.

The major reviews of high stakes accountability studies conclude that the evidence is mixed and provides little scientific foundation for these policies. This has been acknowledged by the chief executive of the Australian Curriculum, Assessment and Reporting Authority in a stunning contradiction of his Minister's claims.

Another stream of studies analyse the impact of choice and competition between schools on student achievement. The weight of evidence from the best designed and most comprehensive of these studies is that increasing choice and competition between schools does not improve student achievement once student and family background characteristics are taken into account.

Greater choice and competition between schools in public education systems does not appear to lead to increased student outcomes. Increased choice and competition from different types of schools within public education systems such as charter schools does not appear to increase student achievement. Nor does competition from private schools appear to increase achievement in public schools. At best, the evidence from these studies is also mixed and therefore does not provide a solid foundation choice and competition policies in education.

Given this evidence, even former advocates of "high stakes" school accountability measures and greater choice and competition between schools, such as former US Assistant Secretaries of Education, Chester Finn and Diane Ravitch, have been forced to admit that the results of these policies have been disappointing and that they have not worked.

1. Introduction

Reporting of school results is intended to provide a means of monitoring progress in school performance and increase the accountability of principals and teachers for school performance. It is claimed that comparisons of school results will engender competition between schools to improve.

It is expected that parents will use this information in ways which will induce schools to improve. Parents may use the information to decide which schools their children will attend, make complaints to schools or to threaten to withdraw their child and enrol in another school. It is thought that the threat of loss of enrolments and funding, as well as public regard, will place pressure on schools to improve performance.

These claims are at the heart of the program for national reporting of school results. The national council of education ministers has stated:

> Through better monitoring of performance at the student, school and system level, educational outcomes can be lifted across all schools. [MCEETYA 2009]

The Prime Minister says it is designed to encourage parents to 'vote with their feet' so that schools will be more responsive to parents.

> ...if some walk with their feet that's exactly what the system is designed to do; that is to make sure that school communities are being responsive to the legitimate high expectations of parents and kids that they get first-class education opportunities at the school level. [Rudd 2008]

The Federal Education Minister, Julia Gillard, says that it will motivate schools to do better.

> I do think transparency of information in and of itself will spur people to do better and they will all want to be seen to be doing better. [Julia Gillard, cited in Ferrari & Nelson 2008]

> Through better monitoring of performance at the student and school level, educational outcomes can be lifted across all schools and all sectors. [Gillard 2009a]

> My School is an important step in the Government's Education Revolution—providing unprecedented transparency and helping drive vital improvements in school education. [Gillard 2010a; see also Gillard 2010b]

Particular emphasis is given to the role of parents in monitoring performance as a way of driving improvement.

> This transparency is critical. To improve schools that are failing their students, we need information. And we want parents to drive change. [Gillard 2009b; see also Gillard 2009c].

This paper reviews the research evidence on the impact of reporting school results on student achievement. The studies reviewed employ quantitative analytical techniques to examine the impact of school reporting and other accountability measures. Three streams of research are reviewed: studies of the effects of school report cards; school accountability studies; and choice and competition studies.

The evidence of research studies is that reporting school results and increasing competition between schools does not improve school performance. There is no conclusive evidence that these policies improve student achievement or reduce achievement gaps. At best, the evidence is mixed and the effects are small. The vast majority of studies are from the United States.

Before proceeding, it should be noted that there are a host of problems in isolating the impact of reporting and other accountability measures on student achievement. For example, many studies examine the impact of accountability programs without distinquishing between the different types of accountability measures used. Studies may confuse the impact of school and non-school factors on student achievement.

Another problem is that studies do not allow for contemporaneous changes in education policy, programs and funding. Changes in curriculum, teaching standards, school organisation and funding may occur within the periods used to assess the impact of accountability measures and are not taken into account in the studies. For example, the introduction of school reporting and high stakes accountability programs in the US was often accompanied by other significant changes, such as the introduction of choice of school, charter schools and changes in funding methods and levels. These and other problems are magnified in the case of cross-country studies.

2. School report card studies

Very few studies have separately assessed the impact of reporting school results, but they all show no significant effect on student achievement. A Brookings Institution study found that school report cards do not have a significant effect on national test scores.

> School report cards and other public reporting strategies appear not to have significant or consistently positive relationships with the NAEP test scores.
> Public reporting is necessary for the implementation of these other policies [rewards and sanctions], but on its own it had no discernible effect on student achievement. [Bishop et.al. 2001: 293, 310]

One of the most frequently cited studies of the impact of school accountability measures on student achievementy analysed the impact in US states that included specific consequences in their accountabilit systems compared to those that only reported school results, without attaching consequences [Hanushek & Raymond 2005]. It found that reporting school results alone had no impact on student achievement.

> The analysis, however, indicates that just reporting results has minimal impact on student performance....
> States that simply provide information through report cards without attaching consequences to performance do not get significantly larger impacts than those with no accountability. [Hanushek & Raymond 2005: 298, 321]

An unpublished study which distinguished the effects of various accountability measures also found that publishing school report cards had no effect on student achievement as measured by national test scores [Harris & Herington 2004, cited in Harris & Herington 2006].

A cross-country study conducted as part of the the Programme for International Student Assessment (PISA) found a small positive effect from reporting schools results [OECD 2007]. However, there are a number of methodological caveats to this effect which give cause to discount such a small effect [OECD 2007: 217-218; Hamilton 2009; Schneider 2009].

3. High stakes accountability studies

Most school accountability studies assess reporting school results along with a range of "high-stakes" accountability measures, which attach consequences to school tests. The consequences applying to schools include monetary rewards for improved performance; sanctions such as closing or re-organizing schools and replacing the principal or teachers; and permitting students in failing schools to enrol elsewhere. High stakes accountability measures also include consequences applying to students. These include grade promotion standards and tests; end-of-course exams; and a high school graduation exam.

Several major studies have compared student achievement between US states with high stakes accountability measures and those with none or low stakes accountability prior to the introduction of the No Child Left Behind (NCLB) legislation which imposed national high stakes accountability on all states. Other studies have examined the impact of NCLB on student achievement. The studies show mixed results, but the weight of evidence is that high stakes accountability programs do not have a significant effect on student achievement.

3.1 Pre-NCLB studies

An extensive review of studies across many US states found that the effects were mixed, inconclusive, and often contradictory [Lee 2007]. These studies found larger or smaller effects depending largely on the statistical methodologies they employed and the data they used.

Another recent review of school accountability studies published in the *Handbook of Research on Education Finance and Policy* found contradictory results [Figlio & Ladd 2008]. It included both cross-state studies and major state-specific studies. It found small or non-existent gains in reading achievement and slightly larger gains in mathematics. It said that it is not a foregone conclusion that school accountability systems will improve school results.

A recent sophisticated meta-analysis of 14 major accountability studies in the US found mixed effects on reading and mathematics achievement [Lee 2008]. The review included cross-state studies which used national test data for reading and/or mathematics. Seven studies favoured states with high stakes testing and reporting, six studies had mixed or insignificant findings, and one study favoured states with low-stakes testing. The average effect on student achievement across these studies of school accountability policies was slightly positive with no significant effect on the racial achievement gap.

There were significant differences in the effects both between and within studies. The effects on reading and mathematics achievement were sometimes modestly negative and sometimes modestly positive. The effects varied among studies depending on the nature, types, and timing of accountability policies. They varied also among subjects, grades, and time periods chosen for the analysis of student achievement. Differences in analytical methods also affected the results. The study concluded:

> This article raises questions about the scientific basis of NCLB and state accountability policy and possible social consequences of the policy on the basis of inconclusive evidence and/or false premises about the policy impact on student achievement. [Lee 2008: 629]

The study also noted that a preliminary review of some recent studies that examined the impact of school and student accountability programs on academic achievement on the basis of state or local data suggest mixed and inconclusive evidence as well [Lee 2008: 628].

The size of the effects in the major studies showing increased student achievement varied. Some studies have found quite large increases in student achievement from accountability consequences for schools [Bishop et.al. 2001; Carnoy & Loeb 2002; Hanushek & Raymond 2005]. A study of high stakes accountability measures in the Chicago public school system found significant increases in student achievement comparable to the effect of class size reductions in the Tennessee STAR project [Jacob 2005]. Others showed a range of results from small to large [Braun 2004] and others small results [Lee & Wong 2004].

The results of two of the most cited and highly regarded of these studies [Carnoy & Loeb 2002; Hanushek & Raymond 2005] mislead about the extent of the gains from accountability

measures. Their measure of achievement gains is based on aggregate state averages and not average gains for individual students. This over-estimates the effect because gains in state averages do not translate into the same level of progress for individual students, whose achievement distributions are much broader than are those of states. Lee [2008] says that most states made very small and insignificant gains that appear larger than they really are when interstate variations in gain scores are used as a criterion to evaluate the effects of policy.

When variation in student level test scores is used as the benchmark, the positive effects of these key studies are only small. For example, the study by Hanushek & Raymond [2005] shows that the introduction of accountability systems with consequences for schools during the 1990s raised 8th grade test scores on the national test scale by an average of only three points, which is very small [see Figlio & Ladd 2008: 177]. Lee's meta-analysis of 14 major studies found that the average increase in student achievement was only three percentage points, the equivalent to two to three months learning [Lee 2008: 625].

The large majority of studies finding positive results used grade-based, data which compares say grade 4 results in one year with grade 4 results in other years. Few analysed the progress of given student cohorts, that is, students who were in grade 4 for one observation period and who were in grade 8 four years later. One study found significant positive results using grade data but the results were reversed when the analysis was done by comparing the progress of particular cohorts of students [Braun 2004]. As students progress through school, there was no difference in achievement trends between states with high-stakes testing and those without. Another study which found some positive effects using grade-based data found no difference between the results in high stakes accountability and low stakes accountability states using cohort data [Nichols et.al. 2006].

The 14 major studies reviewed by Lee provided 62 longitudinal effect estimates, 16 of which were based on cohort data and 46 were grade or age based. The average effect of the cohort estimates was null, while the average grade or age based estimates was slightly positive [Lee 2008]. Lee concludes that "...this contrast indicates that the results of accountability studies are highly sensitive to the choice of analytic method" [623].

Studies that distinguish the impact on student achievement in different grades tend to find that the positive results are larger in the upper primary grades. For example, one frequently cited study of the impact of accountability measures across 50 US states found that students in strong accountability states averaged significantly greater gains on the national 8th grade mathematics test results than students in states with weak accountability systems [Carnoy & Loeb 2002]. However, the 4th grade gains were not as strongly associated with the strength of accountability systems. The Chicago study found that the gains were considerably larger for 8th grade students than those in grades 3 and 6, especially in mathematics [Jacob 2005]. Other studies also found larger gains for students in grade 8 than those in grade 4 [Braun 2004; Nichols et.al. 2006].

Carnoy and Loeb [2002] also found contrasting results on different measures of student achievement. It found large gains in 8th grade achievement in mathematics but no evidence of a positive effect of accountability on student progression through high school. It suggested that one explanation for this contrast could be that the testing programs may be improving test taking skills but not changing factors that influence educational attainment and other outcomes of significance. Nicholls et.al. [2006] found that high stakes accountability was

negatively associated with progression to Year 12 and speculated that it may cause greater numbers of students to drop-out.

Few quantitative studies have examined the effect of high stakes accountability measures on achievement gaps. Carnoy & Loeb [2002] found that the effects were greater for Blacks and Hispanics than for Whites, so that racial achievement gaps narrowed. In contrast, Hanushek & Raymond [2005] found that the Black–White gap widened. Lee & Wong [2004] and Nichols et al. [2006] failed to any significant effects on racial achievement gaps. Lee & Wong also examined changes in the achievement gap among socioeconomic sub-groups of students and found largely insignificant effects on the social-economic achievement gaps.

Several of these studies also looked at the factors that contribute to the positive results from high stakes accountability measures shown in some studies.

There is evidence that observed gains were a result of moving resources from subjects not tested to the subjects that were tested. For example, one study compared trends in mathematics and reading achievement after the introduction of high-stakes testing in Chicago with test score trends in social studies and science, subjects that are not included in the Chicago accountability policy [Jacob 2005]. The gains in mathematics and reading were roughly two to four times larger than gains in science and social studies leading the author to conclude that schools may have shifted resources across subjects.

There is also evidence that a component of any gains from greater accountability may be due to higher rates of exemptions for students in states with strong accountability policies. For example, one study which found greater improvement in US states which had adopted high-stakes accountability measures also found that they were increasingly exempting more students from participating in the national tests [Amrein-Beardsley & Berliner 2003]. It showed that although states with high-stakes measures seemed to outperform those without these measures in fourth-grade mathematics, this difference disappeared when exclusion rates for the tests were taken into account. In contrast, the studies by Carnoy & Loeb [2002] and Hanushek & Raymond [2005] adjusted gain scores for changes in student exclusion rates but this did not lead to significant changes in the estimated effects.

A number of other studies demonstrate that schools respond to accountability pressure by differentially reclassifying low-achieving students as special education so that their scores will not count against the school in accountability systems [for example: Cullen & Reback 2006; Figlio & Getzler 2006; Lemke et.al. 2006]. The Chicago study referred to above also found that schools had increased the proportion of students classified as special education so that they are not required to sit the national tests [Jacob 2005]. The largest increases in special education placements were in low achieving schools. The study also found increased use of grade retention to give students an additional year of learning before moving to the next grade and facing the high-stakes exam.

Another study has showed that that schools differentially suspend students at different points in the testing cycle so as to alter the composition of the testing pool [Figlio 2006]. While schools always tend to assign harsher suspensions to low-performing students than to high-performing students, this gap grows substantially during testing periods for the grades that are tested. Teachers are also more likely to cheat when faced with more accountability pressure [Jacob & Levitt 2003].

Nearly all the key studies showing some positive effects from accountability programs include analysis of student accountability measures along with school measures. Studies have variously included the impact of grade promotion requirements, high school graduation requirements and end-of-course exams on achievement at lower grades [see for example Carnoy & Loeb 2002; Rosenshine 2003; Amrein-Beardsley & Berliner 2003; Braun 2004; Lee & Wong 2004; Jacob 2005]. Of the major studies finding positive results for high stakes accountability, only Bishop et.al. [2001] and Hanushek & Raymond [2005] examined school accountability measures alone.

The inclusion of student (and teacher standards in the case of Lee & Wong 2004) accountability measures is likely to create a bias towards positive results. Studies show that end-of-course exams have significant effects on reading, mathematics, and science achievement [Bishop et.al. 2001; Harris & Herington 2006]. The effects of these exams on student achievement were considerably larger than those of rewards and sanctions for schools.

A later re-working of the data used by Carnoy & Loeb [2002] in their study found that the positive effects of accountability policies on student achievement were largely attributed to the effects of grade promotion exams and graduation exams [Harris & Herington 2006]. The authors concluded that such exams were a "key player" in increased student achievement from more stringent accountability policies.

There is also some suggestion that some part of the results found in studies showing positive results for high stakes accountability measures may be due to unobserved factors, that is, influences on student achievement not included in the analysis. High stakes accountability measures largely refer to test-based measures which were increasingly adopted in the 1990s in the United States. Lee notes that many states also adopted more teacher-based standards during this period and "...looking at only one type of accountability policy may result in an overestimation of the policy effect on student achievement [2008: 617].

To test this Lee added a measure of teacher standards policy to Carnoy & Loeb's regression analysis model of 8th grade mathematics achievement gains. The new analysis found that a large increase in the number of states that required passing exams for new teacher licenses had a significant positive effect on 8th grade mathematics gain scores, whereas the state accountability measures did not.

All this evidence points to a simple conclusion as one leading researcher in the field has stated:

> The bottom line is pretty clear. Accountability has not generated the significant gains in student achievement that policy makers intended... [Ladd 2007: 10]

3.2 NCLB studies

The NCLB legislation was a national school accountability program introduced in 2002. It dramatically expanded the scope accountability for school performance by requiring that states introduce school-accountability systems that applied to all public schools and students in each state. It required annual testing of public school students in reading and mathematics in grades 3 through 8 (and at least once in grades 10-12) and that states publish report cards on schools, both as a whole and for key subgroups, with regard to whether they are making "adequate yearly progress" (AYP). It also required states to introduce "sanctions and

rewards" relevant to every school and based on their AYP status. It mandated sanctions for persistently low-performing schools.

The basic motivation behind NCLB was that publication of detailed information on school performance and linking it to sanctions such as public school choice, staff replacement, and school restructuring would improve student achievement and close achievement gaps.

To date, very few quantitative studies have been published on student achievement under the NCLB. An early study found no increase in student achievement and no decreases in achievement gaps [Lee 2006].

A more recent study found that "the achievement consequences of NCLB are decidedly mixed" and that it has "only made minimal headway in closing achievement gaps" [Dee & Jacob 2009: 36, 37]. The study analysed the impact of NCLB by comparing test-score changes across states that already had school-accountability policies in place prior to NCLB with those that did not.

It found that NCLB generated statistically significant increases in the math achievement of 4th graders of five to seven scale points in states with no prior accountability relative to other states. The gains were concentrated among white and Hispanic students and among students at all levels of performance. It also found a very small increase in 8th grade mathematics, but which was not statistically significant. However, NCLB had no impact on reading achievement in 4th and 8th grade.

The results by race were also mixed. The results for Black, Hispanic and low income students in 4th mathematics increased post-NCLB by more than did those of white students. Hispanic and low income students achieved significant increases in 8th grade mathematics. There were no positive effects by student group in 4th and 8th grade reading. Indeed, NCLB appeared to have reduced the performance of Black and low income students.

The study also tested whether the focus of the NCLB on reading and mathematics caused schools to shift resources at the expense of achievement in other subjects. It found no adverse effect on science achievement.

The only other major study is one of student achievement in seven US states using diagnostic test data provided by a private contractor, not national test data [Ballou & Springer 2009]. It found positive, but small, effects from NCLB.

3.3 Impact on the distribution of student achievement

A few studies of the impact of high stakes accountability regimes have examined their distributional consequences. These studies have found that students just below proficiency standards tend to make greater test score gains, but there is conflicting evidence concerning the effects on students scoring at either the highest or lowest end of the performance spectrum.

Evidence reported on the pre-NCLB accountability system in Texas, for example, suggests that test scores improved most among students at or below the passing threshold, while relatively high-performing students performed worse than expected [Reback 2008]. Schools have incentives to neglect their high achieving students to focus on low achievers, but the extent of this differed between subjects. The evidence from this study was that schools

responded to incentives related to mathematics in ways that increased the performance of low performing students with only small adverse effects on higher achieving students. In reading, by contrast, students who had a moderate to strong probability of passing the exam appeared to be harmed.

Some recent studies have also examined the impact of NCLB on the distribution of student test results.

The first of these studies found that NCLB's threat of sanctions are positively correlated with test score gains by low performing students in failing schools, and that greater than expected test score gains by low performing students did not occur at the expense of high performing students in failing schools [Springer 2008]. A study using test results in Washington state found that NCLB provides incentives to schools to focus students at the margin of proficiency thresholds at the expense of low and high achieving students [Krieg 2008]. Students at the lower and upper ends of the achievement distribution gained less at schools facing sanctions. In contrast, a study of test results in seven US states found little or no evidence that schools facing accountability pressure neglect their high achieving students to focus on low achievers [Ballou & Springer 2009].

A recent study of North Carolina student test data analysed the distribution of results in terms of both the level of achievement and that gains in achievement in reading and mathematics [Ladd & Lauen 2009]. It found that the distributional effects differ depending on whether the system holds schools accountable for improvement in student achievement or for the level of achievement. Under both approaches to accountability, students below the proficiency standard typically benefit relative to those just above the standard, although the pattern is more consistent for mathematics than for reading. However, accountability based on the level of student achievement appears to reduce the reading achievement of students above the proficiency threshold. No such negative effects emerge with respect to the accountability measure based on gains in student achievement.

The study found little evidence that students at the bottom of the achievement distribution were ignored under either accountability system. The lowest achieving students had positive, although sometimes zero, achievement gains.

A newly published study compared mathematics and reading test scores of Chicago Public School students in 5^{th} and 6^{th} grade before and after the implementation of two separate high-stakes accountability systems [Neal & Schanzenbach 2010]. It found significant increases in mathematics and reading test scores among those around the accountability system's proficiency threshold, while the performance of low-performing students was the same or lower. Effects on the achievement of high-performing students were mixed. The study concluded that schools with a significant number of low achieving students face a strong incentive to shift attention away from these students and toward students near proficiency thresholds:

> NCLB provides relatively weak incentives to devote extra attention to students who have no realistic chance of becoming proficient in the near term or students who are already proficient....
> Students who have no realistic chance of becoming proficient in the near term appear to gain little from the introduction of these systems. [Neal & Schanzenback 2010: 264, 277]

3.4 Cross-country studies

A few studies have utilized international test data to analyse the impact of accountability measures on student achievement.

One such study was reported as part of the PISA 2006 [OECD 2007]. The initial modellng results reported in the PISA study show that students in schools that publicly release their performance results performed 14.7 score points better than students in schools that did not. This difference amounts to a little less than six months of learning, as 37 score points on the PISA science scale is equivalent to about one year of learning growth. However, when demographic and socio-economic background factors were taken into account this difference due to reporting school results was reduced to 6.6 score points, or about two months learning growth. Further, when the joint impact of school and system resources, practices and policies was taken into account to eliminate overlapping effects, the unique difference in student achievement due to reporting school results was reduced to 3.5 scale points, which is very small.

There was enormous variation in the practices of the highest performing countries in reporting school results. Among the five highest achieving countries, the percentage of students who attend schools that report school results ranges from 4% in Finland, 11% in Japan and 17% in Korea to 64% in Canada and 60% in Australia [Table 5.22]. The average across all five countries was 31% and the OECD average was 38%. As the former head of the U.S. Department of Education's National Center for Education Statistics, Mark Schneider, says:

> Best practices might suggest *not* posting results, but the PISA report argues that accountability matters and that posting results is a way of improving performance. [Schneider 2009: 11]

It should also be noted that the OECD issued a number of methodological caveats to the study as cautions in drawing prescriptive policy conclusions [OECD 2007: 215-216]. Further, the design of PISA and its data limit the extent to which any strong conclusions can be made on the basis of this type of cross-national analyses [Hamilton 2009]. There are a number of issues relating to the interpretation of the PISA survey questionnaire across countries, use of a study design that does not support causal inferences, data for a single year and reliance on very small samples of school principals as the source of information. As a result:

> ....the presentation of results and the discussions of implications could lead readers to make conclusions that are not warranted based on the data and analyses used. [Hamilton 2009: 7]

An example of these issues is the use of multi-level regression models using cross-sectional data which do not support the kinds of inferences made because a number of unmeasured factors could influence the magnitude and even the direction of an observed relationship between achievement and a school-based characteristic such as reporting school results. This and the other issues “raise doubts about the extent to which PISA can be used to support causal inferences about education policies and practices” [Hamilton 2009: 17].

Other cross-country accountability studies have focused almost exclusively on one specific accountability device, namely, external exit exams at the end of secondary school. The evidence from several of these studies is that student achievement is significantly higher in countries that have external exit exams than in countries without external exit-exam systems [Woessmann et.al. 2008; Woessmann 2007; Hanushek & Woessmann 2010]. For example, studies using the PISA 2000 and 2003 data suggests that students perform better in countries

which have external exit exams, after allowing for other factors such as student characteristics, family background, school funding and school organization. [Fuchs & Woessmann 2007; Woessmann et.al. 2008].

One study published by the OECD shows that student achievement is higher in schools which use assessments to compare themselves to district or national performance [Woessmann et.al. 2008]. However, the effect is quite small. A complementary study found no differential effect on students from different socio-economic backgrounds [Schutz et.al. 2008].

It should be noted, however, that the methodological problems identified with the PISA 2006 report also apply to these other studies prepared for the OECD relating school characteristics to PISA student achievement data.

3.5 Australian Government claims

The Australian Treasury and the Department of Education, Employment and Workplace Relations (DEEWR) have cited some cross-country studies in support of the claim that the publication of school test results tends to improve school performance [Commonwealth of Australia 2008; Bruniges 2009].

One of the papers cited by the Treasury is by well-known academics who have been strong supporters of school accountability systems [Hanushek & Woesmann 2007]. The paper is a broad review of the role of school improvement in economic development and reviews the evidence on several policy approaches to school improvement, including school accountability.

Its review of the evidence on school accountability approaches is brief and selective. It cites three pre-NCLB studies reviewed above, all of which show increased student achievement associated with high stakes accountability systems. The paper does not mention other studies that find mixed results or no benefits. Moreover, each of the studies cited include a range of both school and student accountability measures and, as discussed above, it is student accountabilty measures which appear to contribute most to the positive results in these studies. Indeed, one of the cited studies [Hanushek & Raymond 2005] found that reporting school results alone has no effect on student achievement and directly contradicts the Treasury claim.

Another of the cited studies also found that part of the gains in student achievement is the result of schools responding strategically to the incentives created by high stakes testing [Jacob 2005]. It found evidence that schools manipulate results in a variety of ways, including teaching to the test, increasing special education placements, pre-emptively retaining students and substituting resources away from low-stakes subjects like science and social studies. It concluded:

> Overall, these results suggest that high-stakes testing has the potential to improve student learning, but may also lead to some undesired strategic responses on the part of teachers including a narrowing of teaching to focus on the set of skills emphasized on the high-stakes test. [763]

Another study cited is the most recent PISA report [OECD 2007] reviewed above. The report compares PISA results across countries according to whether they have school reporting or not after taking into account a range of other school factors and the socio-economic background of students.

The Deputy Secretary of DEEWR, in particular, cites the results of this study in a very misleading way [Bruniges 2009]. She only cites the results of the initial modelling of the relationship between reporting school results and student achievement and fails to cite the results of the full modelling which takes account of other school factors which may overlap with the effects of school reporting. She also fails to cite the result of adjustments for the demographic and socio-economic background of students. Given that all these results are reported in the same chapter of the report it is difficult not to conclude that the results were cited in a way designed to deceive and mislead.

The chief executive of the Australian Curriculum, Assessment and Reporting Authority, Peter Hill, admits there is little evidence to support his Minister's claims. Hill told a *Daily Telegraph* forum on league tables in Sydney last year that the evidence is mixed [Parker 2009]. He said that while some studies show that student achievement does improve, many studies show little or no impact. He also said that the effects are generally small, and he noted that there was no evidence that racial inequalities were reduced by public reporting. This is a stunning admission which contradicts government claims.

4. Choice and competition studies

The other relevant stream of research studies analyses the impact of greater choice and competition between schools. This stream is relevant because the expectation is that reporting school results informs school choice and creates greater competition between schools which leads to improved student achievement.

There is a huge literature on the impact of choice and competition on student achievement and achievement gaps. There are a range of findings, from significantly positive to negative. The weight of evidence from the best designed and most comprehensive research studies around the world is that increasing choice and competition between schools does not improve student achievement once student and family background characteristics are taken into account.

It is only possible here to provide a brief overview of the major conclusions from the more recent studies [see Cobbold 2007 for a more extensive review].

First, greater choice and competition within public education systems does not appear to lead to increased student outcomes. At best, the evidence is mixed.

A recent and very sophisticated study of English primary schools by researchers at the London School of Economics concluded that there is "no systematic impact of choice and competition on pupil performance" and that "pupil achievement is generally unrelated to the competitive pressures a school faces" [Gibbons et.al. 2008: 939, 942]. It also concluded from its review of the international literature on the effects of choice and competition that "the existing evidence is mixed, and at best offers a shaky foundation for policy" [913].

Similar findings have been made in US studies. For example, a study of the impact of the open enrolment scheme for Chicago public high schools on graduation rates found that that systemic choice within a public school district does not seem to benefit those who participate [Cullen et.al. 2005]. It concluded that:

> This casts doubt on the power of this form of choice to improve educational outcomes, or to serve as an efficient form of discipline for low quality neighbourhood schools. [755]

The same authors have also analysed the results of lotteries used to allocate students to high demand Chicago high schools [Cullen et.al. 2006]. In comparing lottery winners and losers, it found no evidence that winning a lottery provided any benefit on a wide variety of traditional achievement measures, including standardized test scores, graduation and attendance rates. In other words, there was no positive impact on student achievement of public school choice. In some cases, it found that lottery winners were worse off.

A further study has analysed the results of lotteries used to allocate students to high demand Chicago elementary schools with similar results [Cullen & Jacob 2009]. Similar results have also been found in school districts in North Carolina [Hastings et.al. 2006, 2007].

Second, increased choice and competition from different types of schools within public education systems does not appear to increase student achievement.

An extensive review of available research studies on various forms of school choice, including charter schools, magnet schools and voucher schools, was published by the Federal Reserve Bank of Chicago [Rouse & Barrow 2008]. It concluded that students who exercise these various forms of choice do not experience achievement gains and that school choice does not induce public schools to improve their performance.

The most rigorous and comprehensive study of student achievement in charter schools in the United States to date found that charter school results were worse than or no better than those of traditional public schools [CREDO 2009]. The study compared the performance of charter schools against their local public school alternatives. It found that the gains in maths results for nearly half of all charter schools (46%) are no different from those in comparable traditional public schools while over one third (37%) of charter schools have significantly worse results. Only 17% of charter schools have significantly higher maths results than students in comparable traditional public schools.

The study also analysed the aggregate impact of charter schools on student performance using a nationally pooled data set. On average, the learning growth of charter school students was lower than their traditional public school peers, although the absolute differences are quite small. The gains in reading for charter school students were only slightly below that of traditional public school students while the gains in maths were significantly less for charter school students.

The study concluded:

> ...this study reveals in unmistakable terms that, in the aggregate, charter students are not faring as well as their TPS counterparts. Further, tremendous variation in academic quality among charters is the norm, not the exception. The problem of quality is the most pressing issue that charter schools and their supporters face. [6]

These results have strengthened the body of research evidence showing average charter performance to be equal to, or lower than, the performance of traditional schools. Similar findings were made in another study of charter schools across eight US states by the RAND Corporation [Zimmer et.al. 2009]. It found that student achievement in charter schools does not differ substantially from those of traditional public schools and that competition from charter schools does not increase student achievement in nearby traditional public schools.

A review of major studies of charter schools published in the *Handbook of Research in Education Finance and Policy* concluded:

> Research to date provides little evidence that the benefits envisioned in the original conception of charter schools – organizational and educational innovation, improved student achievement and efficiency – have materialized…Convincing evaluation of student achievement effects are now in from five different states. In none of these states have charter schools, on average, had large or unequivocally positive effects on student achievement. [Bifulco & Bulkley 2008: 440].

Third, competition from private schools does not appear to increase achievement in public schools. For example, a recent study by the London School of Economics examined the private sector 'competitive threat' on the performance of public sector schools and found no evidence that a higher concentration of private schools improves the performance of neighbouring public sector schools in England [Gibbons & Silva 2008]. A study of religious secondary schools within the English public sector by the Institute of Education at the University of London found no evidence that competition from religious schools raises overall area-wide educational performance [Allen & Vignoles 2009].

Voucher programs whereby low income students are given government funding to attend private schools are also seen by many as a way to increase student achievement through competition. The oldest and largest voucher program operating in the United States is in the city of Milwaukee. Recent evaluations of the program have found no statistically significant difference in achievement progress between students participating in the program and students in Milwaukee public schools [Witte et.al. 2009; Witte et.al. 2010].

All this evidence has lead Professor of Economics at the University of Chicago and the co-author of the best-selling *Freakonomics*, Steven Levitt, to say of school choice that, "the theory sounds great, but evidence confirming it has been hard to find" [Levitt 2007]. Another recent review of the literature concludes that: "All in all, it is fair to say that the international and US evidence is voluminous, but mixed in its findings" [Silva 2009: 17].

In the light of all this, it is interesting to note that one of the most vigorous advocates of increasing school choice, competition, charter schools, vouchers and accountability measures in education over the past 20 or more years recently acknowledged that it hasn't worked. Last month, Chester Finn, a former US Assistant Secretary of Education under Ronald Reagan, said:

> ....despite all the reforming, U.S. [test] scores have remained essentially flat, graduation rates have remained essentially flat, and our international rankings have remained essentially flat. You can find some upward blips but you can also find downward blips. Big picture, over 25 years, is flat, flat, flat. In other words, all the reforming has yielded little or nothing by way of stronger outcomes. [Finn 2009]

Another former US Assistant Secretary of Education and former advocate of choice and competition in schooling, Diane Ravitch, also says that it has not worked:

> Our schools will not improve if we entrust them to the magical powers of the market. [Ravitch 2010a: 227]

> Choice and accountability, I believed, would offer a chance to poor children to escape failing schools. Testing and accountability, I thought, would cast sunshine on low-performing schools and lead to improvement...
> Today there is empirical evidence, and it shows clearly that choice, competition and accountability as education reform levers are not working. [Ravitch 2010b]

Indeed, the evidence is that increased competition as a result of school comparisons and rankings appears to more lead to standardization rather than innovation in classroom practice and curriculum. A recent OECD study summarising evidence on innovations in more market-driven education systems in over 20 countries found that competition between schools tends to promote uniformity rather than innovation and an emphasis on marketing rather than educational change [Lubienski 2009].

5. Conclusions

The first major conclusion from the above analysis is that there is no evidence that reporting school results increases student achievement.

A second conclusion is that studies of increased "high stakes" accountability measures for school performance show mixed results. The most comprehensive review of the research studies on high stakes accountability systems operating prior to the NCLB found that seven studies concluded that high stakes testing and reporting increased student achievement, six studies had mixed or insignificant findings, and one study favoured low stakes testing. Other reviews of research studies also conclude that the evidence is mixed.

There are significant differences in the effects both between and within studies. The effects on reading and mathematics achievement were sometimes modestly negative and sometimes modestly positive. The effects varied among studies depending on the nature, types, and timing of accountability policies. They varied also among subjects, grades, and time periods chosen for the analysis of student achievement. Differences in analytical methods also affected the results.

Other evidence on the impact of high stakes accountability measures on student achievement comes from more recent studies of the impact of the No Child Left Behind legislation which imposed national requirements for school accountability in the United States. The evidence from these studies is also mixed. Studies which use international test results to examine the impact of school accountability measures have significant methodological issues.

A third significant conclusion is that most highly regarded of the studies showing positive effects use a misleading benchmark of effect size which over-estimates the impact on student achievement. Their measure of achievement gains is based on aggregate state averages and not average gains for individual students. This over-estimates the effect because gains in state averages do not translate into the same level of progress for individual students, whose achievement distributions are much broader than are those of states. When measured against the average gains for students, the increase in student achievement in these studies is very small.

A further conclusion is that the major source of gains in student achievement found in some studies is from measures targeted directly at students – so-called student accountability measures – rather than school accountability. Nearly all the key studies showing some positive effects from accountability programs include analysis of student accountability measures along with school measures. Studies have variously included the impact of grade promotion requirements, high school graduation requirements and end-of-course exams on achievement at lower grades. These factors contributed significantly to the small positive results of these studies and there is some evidence that their effect is larger than the effects of rewards and sanctions relating to school performance.

The fifth major conclusion is that there is evidence that the positive results in some studies may be due, at least in part, to schools manipulating their results in various ways. Studies show that schools respond to high stakes accountability tests by moving resources from subjects not tested to the subjects that were tested, re-classifying students as special education students so they can be exempted from high stakes tests, suspending low achieving students during the testing cycle and by cheating.

The few studies that have examined the distributional consequences of high stakes accountability regimes found that students just below proficiency standards tend to make greater test score gains as schools tend to concentrate resources on improving the results of this group. There is conflicting evidence concerning the effects on students scoring at either the highest or lowest end of the performance spectrum.

Another stream of studies analyse the impact of choice and competition between schools on student achievement. The weight of evidence from the best designed and most comprehensive of these studies around is that increasing choice and competition between schools does not improve student achievement once student and family background characteristics are taken into account.

Greater choice and competition within public education systems does not appear to lead to increased student outcomes. Increased choice and competition from different types of schools within public education systems such as charter schools does not appear to increase student achievement. Nor does competition from private schools appear to increase achievement in public schools. At best, the evidence from these studies is mixed and therefore does not provide a solid foundation choice and competition policies in education.

References

Allen, R. & Vignoles, A. 2009. Can School Competition Improve Standards? The Case of Faith Schools in England? Paper presented to the Annual Conference of the Royal Economic Society, April. Available at: http://repec.ioe.ac.uk/REPEc/pdf/qsswp0904.pdf

Amrein-Beardsley, A. & Berliner, D. 2003. Re-analysis of NEAP Math and Reading Scores in States With and Without High-stakes Tests: Response to Rosenshine. *Education Policy Analysis Archives*, 11 (25).

Ballou, D. & Springer, M. G. 2009. Achievement Trade-offs and No Child Left Behind. Paper presented to the *NCLB: Emerging Findings Research Conference* at the Urban Institute, Washington, DC, 12 August 2009. Available at: http://www.caldercenter.org/nclbconference.cfm

Bifulco, R. & Bulkley, K. 2008. Charter Schools. In Ladd, H.F. & Fiske, E.B. (eds.). *Handbook of Research on Education Finance and Policy*. London: Routledge.

Bishop, J.; Mane, F.; Bishop, M. & Moriarity, J. 2001. The Role of End-of-Course Exams and Minimal Competency Exams in Standards-Based Reforms. In Diane Ravitch (ed.), *Brookings Papers in Educational Policy*, Brookings Institution, Washington DC: 267-344.

Braun, H. 2004. Reconsidering the Impact of High-Stakes Testing. *Education Policy Analysis Archives*, 12 (1), January.

Bruniges, M. 2009. Opening Address. Australasian Curriculum, Assessment and Certification Authorities Conference. University of Western Australia, Perth, 13 July.

Carnoy, M. & Loeb, S. 2002. Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24 (4): 305–331.

Center for Research on Education Outcomes (CREDO) 2009. Multiple Choice: Charter School Performance in 16 States. Stanford University, Stanford, CA, June. Available at: http://credo.stanford.edu/

Cobbold, T. 2007. The Great School Fraud: Howard Government School Education Policy 1996-2000. Available at: http://www.saveourschools.com.

Commonwealth of Australia 2008. Boosting Australia's Productive Capacity: Infrastructure Skills. *Statement No. 4, Budget Paper No. 1: Budget Strategy and Outlook 2008-09.* Available at: http://www.budget.gov.au/2008-09/content/bp1/html/index.htm

Cullen, J & Jacob, B. 2009. Is Gaining Access to Selective Elementary Schools Gaining Ground? Evidence from Randomized Lotteries. In Jonathan Gruber (ed.), *An Economics Perspective on the Problems of Disadvantaged Youth*. National Bureau of Economic Research, University of Chicago Press, Chicago, 43-84.

Cullen, J.; Jacob, B. & Levitt, S. 2005. The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools. *Journal of Public Economics*, 89 (5-6): 729-60.

Cullen, J.; Jacob, B. & Levitt, S. 2006. The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries. *Econometrica* 74 (5): 1191–1230.

Cullen, J. B. & Reback, R 2006. Tinkering Toward Accolades: School Gaming under a Performance Accountability System. Working Paper 12286, National Bureau of Economic Research, Cambridge, June. Available at: http://www.nber.org/papers/w12286

Dee, T. & Jacob, B. 2009. The Impact of No Child Left Behind on Student Achievement. Working Paper 15531, National Bureau of Economic Research, Cambridge, November. Available at: http://www.nber.org/papers/w15531

Ferrari, J. & Nelson, D. 2008. Gillard Admits School Choice is 'Guesswork'. *The Australian*, 23-24 August.

Figlio, D. 2006. Testing, Crime and Punishment. *Journal of Public Economics*, 90: 837-851.

Figlio, D & Getzler, L 2006. Accountability, Ability and Disability: Gaming the System. In Timothy J. Gronberg and Dennis W. Jansen (eds.) Improving School Accountability: Check-Ups or Choice, *Advances in Applied Microeconomics*, 14: 35–49.

Figlio, D. & Ladd, H. 2008. School Accountability and Student Achievement. In Ladd, H.F. & Fiske, E.B. (eds.). *Handbook of Research on Education Finance and Policy*. London: Routledge.

Finn, C. 2009. Is it Time to Throw in the Towel on School Reform? Speech to Jones Graduate School of Business, Rice University, Houston, Texas, 9 September. Available at: http://edexcellence.net/doc/200909_CheckerFinnSpeech.pdf

Fuchs, T. & Woessmann, L. 2007. What Accounts for International Differences in Student Performance? A Re-examination using PISA Data. *Empirical Economics* 32 (2-3): 433-464.

Gibbons, S.; Machin, S. & Silva, O. 2008. Choice, Competition and Pupil Achievement. *Journal of the European Economic Association* 6 (4): 912–947.

Gibbons, S. & Silva, O. 2008. Urban Density and Pupil Achievement. *Journal of Urban Economics*, 63: 631-650.

Gillard, J. 2009a. Rudd Government Committed to Greater Transparency. Media Release, 26 May.

Gillard, J. 2009b. A New Progressive Consensus for Australian Schools. Presentation to the Brookings Institution, Washington DC, 19 June.

Gillard, J. 2009c. Parents Deserve More Information About Schools. *The Punch*, 28 July.

Gillard, J. 2010a. My School Website to go Live Tomorrow. Media Release, 27 January.

Gillard, J. 2010b. My School Website Launched. Media Release, 28 January.

Hamilton L. S. 2009. Using PISA Data to Measure School and System Characteristics and Their Relationships with Student Performance in Science. Paper presented to National Centre for Education Statistics research conference on the Programme for International Student Assessment: What We Can Learn From PISA, Washington DC, 2 June. Available at: https://edsurveys.rti.org/PISA/

Hanushek, E. & Raymond, M. 2005. Does School Accountability Lead to Improved School Performance? *Journal of Policy Analysis and Management*, 24 (2): 297-329.

Hanushek, E. & Woessmann, L. 2007. The Role of School Improvement in Economic Development. Working Paper 12832, National Bureau of Economic Research, Cambridge, MA, January. Available at: http://www.nber.org/papers/w12832

Hanushek, E. & Woessmann, L. 2010. The Economics of International Differences in Educational Achievement. Discussion Paper No. 4925, Institute for the Study of Labour, Bonn, April. Available at: http://ftp.iza.org/dp4925.pdf

Harris, D. & Herrington, C. 2006. Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-century. *American Journal of Education*, 112 (2): 209-238.

Hastings, J. S.; Kane, T. & Staiger, S. 2006. Parental Preferences and School Competition: Evidence from a Public School Choice Program. Working Paper 11805, National Bureau of Economic Research, Cambridge, MA.

Hastings, J. S.; Kane, T. & Staiger, S. 2007. Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery. Working Paper 12145, National Bureau of Economic Research, Cambridge, MA.

Jacob, B. A. 2005. Accountability, Incentives and Behaviour: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics* 89 (5-6): 761-796.

Jacob, B. A. & Levitt, S. 2003. Rotten Apples: An Investigation of the Prevalence and Predictions of Teacher Cheating. *Quarterly Journal of Economics*, 118 (3): 843-877.

Krieg, J.M. 2008. Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act. *Education Finance and Policy*, 3 (2):250-281.

Ladd, H. F. 2007. Holding Schools Accountable Revisited. Spencer Foundation Lecture in Education Policy and Management, Association for Public Policy Analysis and Management, Washington DC, 8 November.

Ladd, H. F. & Lauen, D. L. 2010. Status versus Growth: The Distributional Effects of School Accountability Policies. *Journal of Policy Analysis and Management*, 29 (3): 426-450.

Lee, J. 2006. Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An In-depth Look into National and State Reading and Math Outcome Trends. The Civil Rights Project, Harvard University, Cambridge, Mass., June. Available at: http://www.civilrightsproject.ucla.edu/research/esea/esea_gen.php

Lee, J. 2007. *The Testing Gap: Scientific Trials of Test-driven School Accountability Systems for Excellence and Equity*. Charlotte, NC: Information Age Publishing.

Lee, J. 2008. Is Test-Driven External Accountability Effective? *Review of Educational Research*, 78 (3): 608-644.

Lee, J. & Wong, K.K. 2004. The Impact of Accountability on Racial and Socioeconomic Equity: Considering Both School Resources and Achievement Outcomes. *American Educational Research Journal* 41 (4): 797–832.

Lemke, R.J; Hoerandner, C.M & McMahon, R.J 2006. Student Assessments, Non-test Takers, and School Accountability. *Education Economics*, 14 (2): 235-250.

Levitt, S. 2007. More Evidence on the (Lack of) Impact of School Choice. *New York Times*, 4 October.

Lubienski, C. 2009. Do Quasi-Markets Foster Innovation in Education? A Comparative Perspective. Education Working Paper No. 25, Directorate for Education, Organisation for Economic Co-operation and Development, Paris, August. Available at: http://www.oecd.org/dataoecd/41/6/43415929.pdf

Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA] 2009. Communique, 17 April.

Neal, D. & Schanzenbach, D. 2010. Left Behind By Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 92 (2): 263–283.

Nichols, S. L., Glass, G. V, & Berliner, D. C. 2006. High-stakes Testing and Student Achievement: Does Accountability Pressure Increase Student Learning? *Education Policy Analysis Archives*, 14 (1), January.

Organisation for Economic Co-operation and Development (OECD) 2007. *Science Compentencies for Tomorrow's World*. OECD, Paris.

Parker, M. 2009. The Minefield that is League Tables. *The Daily Telegraph*, 28 October.

Ravitch, D. 2010a. *The Death and Life of the Great American School System*, Basic Books, New York.

Ravitch, D. 2010b. The Big Idea – It's Bad Education Policy. *Los Angeles Times*, 14 March.

Reback, R. 2008. Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics*, 92(5-6): 1394-1415.

Rosenshine, B. 2003. High-stakes Testing: Another Analysis. *Education Policy Analysis Archives*, 11(24).

Rouse, C.E. & Barrow, L. 2008. School Vouchers and Student Achievement: Recent Evidence, Remaining Questions. Working Paper 2008-08, Federal Reserve Bank of Chicago. Available at:

http://www.chicagofed.org/economic_research_and_data/publication_display.cfm?Publication=6&year=2006%20AND%202009 .

Rudd, K. 2008. Questions and Answers. National Press Club, Canberra, 27 August. Available at: http://www.pm.gov.au/node/5620

Schneider, M. 2009. International Benchmarking. Paper presented to National Centre for Education Statistics research conference on the Programme for International Student Assessment: What We Can Learn From PISA, Washington DC, 2 June. Available at: https://edsurveys.rti.org/PISA/

Schutz, G.; West M.R. & Woessmann, L. 2008. School Accountability, Autonomy, Choice, and the Equity of Student Achievement: International Evidence from PISA 2003. Education Working Paper No. 14, EDU/WKP(2007)9/REV 1, Directorate of Education, OECD, Paris, 14 November. Available at: http://www.olis.oecd.org/olis/2007doc.nsf/LinkTo/NT00008A2E/$FILE/JT03255311.PDF

Silva, O. 2009. Some Remarks on the Effectiveness of Primary Education Interventions. Policy Paper No. 5, Institute for the Study of Labour, Bonn, April. Available at: http://ftp.iza.org/pp5.pdf

Springer, M. G. 2008. The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains. *Economics of Education Review* 27 (5): 556-563.

Witte, J. F.; Wolf, P. J.; Cowen, J. M.; Fleming, D. J. and Lucas-McLean, J. 2009. *The MPCP Longitudinal Educational Growth Study: Second Year Report*, SCDP Milwaukee Evaluation Report #10, School Choice Demonstration Project, University of Arkansas, Fayetteville, AR, March. Available at: http://www.uark.edu/ua/der/SCDP/Milwaukee_Research.html

Witte, J. F.; Cowen, J. M.; Fleming, D. J.; Wolf, P. J.; Congdon, M.R. and Lucas-McLean, J. 2010. *The MPCP Longitudinal Educational Growth Study: Third Year Report*, SCDP Milwaukee Evaluation Report #15, School Choice Demonstration Project, University of Arkansas, Fayetteville, AR, April. Available at: http://www.uark.edu/ua/der/SCDP/Milwaukee_Research.html

Woessmann, L. 2007. International Evidence on School Competition, Autonomy and Accountability: A Review. *Peabody Journal of Education 82 (2-3)*: 473-497.

Woessmann, L.; Lüdemann, E.; Schütz, G. & West M.R. 2008. School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003. Education Working Paper No. 13, EDU/WKP(2007)8/REV1, Directorate of Education, OECD, Paris, 14 November. Available at: http://www.olis.oecd.org/olis/2007doc.nsf/ENGDIRCORPLOOK/NT00008A32/$FILE/JT03255304.PDF

Zimmer, R.; Gill, B.; Booker, K.; Lavertu, S.; Sass, T. R. & Witte, J. 2009. *Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition*. RAND Corporation, Santa Monica.

Publication of School Results and League Tables Harm Education

Submission to the Senate Education Committee Inquiry into the Administration and Reporting of NAPLAN

Part 2

SAVE OUR SCHOOLS

June 2010

http://www.saveourschools.com.au

Table of Contents

Key Points

1. Far from improving education, publication of school results and league tables are likely to significantly harm education. Overseas experience shows that they:
   - Narrow the curriculum;
   - Distort teaching practice;
   - Disadvantage low and high achieving students;
   - Unfairly stigmatise low achieving students;
   - Make it more difficult for low performing schools to retain high quality teachers;
   - Discourage co-operation and collaboration between schools and teachers; and
   - Increase social segregation and inequity in education.

2. The curriculum is narrowed in at least two ways:
   - More time is devoted to the tested subjects of literacy and maths while untested subjects such as science, history, social studies, languages, arts and music, physical education and health receive much less time. Even recess gets cut;
   - In the subjects that are tested, greater emphasis is given to the areas that are most conducive to testing by multiple-choice questions and there is less teaching of more complex skills.

3. Teaching is distorted because schools and teachers tend to respond to pressure created by publication of school results and league tables by focusing more on teaching test taking skills and practicing for tests.

4. Publishing school results and league tables can undermine efforts to improve the quality of teaching because they turn the notion of a "good teacher" into one who increases test scores.

5. Schools tend to concentrate on improving the results of students who are on the border of accepted benchmarks at the expense of both high and low achieving students.

6. Publication of the results of individual schools and public rankings of schools may unfairly stigmatise and humiliate some students and alienate them from schooling.

7. Published school results and league tables may be used as a job guide and make it more difficult for schools with low results to retain and attract quality teachers and principals.

8. Publication of school results and competition for league table rankings can reduce collaboration between schools and between teachers within schools.

9. Publication of school results and league tables tends to increase socio-economic and ethnic segregation between schools which exacerbates inequity in education because:
   - Student learning needs in some schools increase without proportionate increases in resources to meet those needs;
   - Increasing concentrations of students from low socio-economic status families in some schools tend to lead to lower overall outcomes.

10. There is evidence of some of these effects already in Australia after only one year of publishing school results.

Summary

Far from improving education, publication of school results and league tables are likely to significantly harm education. Overseas experience shows that they:

- Narrow the curriculum;
- Distort teaching practice;
- Disadvantage low and high achieving students;
- Unfairly stigmatise low achieving students;
- Make it more difficult for low performing schools to retain high quality teachers;
- Discourage co-operation and collaboration between schools and teachers; and
- Increase social segregation and inequity in education.

There is evidence of some of these effects already in Australia after only one year of publishing school results.

Narrowing of the curriculum

Publication of school results and league tables restrict student learning because they narrow the curriculum and teaching. Students receive a less rounded education.

There are two major aspects of narrowing the curriculum. First, more time and resources are devoted to the tested subjects while other subjects not tested are neglected. Overseas evidence shows that schools direct more resources into the tested subjects of literacy and maths while untested subjects such as science, history, social studies, languages, arts and music, physical education and health receive much less time. Even recess gets cut.

Second, within those subjects that are tested, there is greater focus on the skills that are most conducive to testing by multiple-choice questions and less teaching of more complex thinking and writing skills.

There is already some evidence that publishing NAPLAN school results on My School is narrowing the curriculum in secondary schools in Australia. A survey conducted by the Australian Secondary Principals Association found that 33% of principals said that publication of school results had reduced the breadth of curriculum in their school.

Distorting teaching practices

Teaching practice also tends to change under the pressure of publishing school results and the pressure to improve league table rankings.

There is extensive evidence from overseas that schools and teachers tend to respond to publication of school results and league tables by focusing more on teaching test taking skills and practicing for tests at the expense of deeper learning experiences that develop analytical skills and greater understanding. League tables turn classrooms into test preparation factories. Weeks and months are devoted to test preparation.

There is considerable evidence that publication of NAPLAN school results on the My School website and publication of school league tables is causing schools to spend large amounts of time on test practice in class to the detriment of other areas of the curriculum. A survey conducted by the Australian Secondary Principals Association found that 65% of schools reported that they had increased the time spent in class on preparation for the NAPLAN tests in 2010 and 70% said they had increased the time spent on practising tests. A report

published by the Australian Primary Principals Association said large amounts of valuable instructional time was taken up by coaching and practising tests.

Many teachers now refer to NAPLAN as “napalm” because “it kills everything in the classroom”.

There is also evidence that school administrators in some states have pressured principals and schools to devote more time to practising for tests.

This focus on test preparation and raising test scores may undermine efforts to improve the quality of teaching in schools. Some studies show that the idea of a “good teacher” changes in schools under the pressure to lift school results. A “good teacher” becomes someone who increases test scores rather than someone who guides students to deeper and wider learning experiences and encourages self-motivated learning.

Disadvantages low and high achieving students

The pressure to improve school results can also create incentives for schools to ignore students at the lower and upper ends of the student achievement scales. Several English and US studies have found that schools concentrate on improving the results of students who are on the border of accepted benchmarks and neglect the lowest and highest achievers.

Improving the results of students just below benchmarks is seen as the most efficient way to increase a school’s average score or the proportion of students achieving a benchmark. However, while this strategy gives schools a better chance of improving their ranking, it may lead to worse outcomes for low and high performing students.

In addition, publishing outcomes of individual schools and public rankings of schools often unfairly stigmatises and humiliates schools, teachers, students and their families. Students and teachers in particular years may be easily identifiable as the “culprits” when a school gets a low ranking, especially in small schools, of which there are many in rural areas of Australia. Students who are humiliated for their lower learning accomplishments are unlikely to respond positively in their future learning.

Harder for low performing schools to retain quality teachers

League tables of school results serve as a job guide for teachers to apply to highly ranked schools with fewer learning and behavioural problems. This means that low ranked schools often end up with the least qualified, least experienced teachers.

Discourages collaboration and co-operation

Collaboration is generally seen as an important way to spread innovative approaches and good teaching practice both between and within schools. Publication of school results and competition for rankings can reduce collaboration between schools and teachers and slow the dissemination of best practice.

Schools will be reluctant to share successful practices with other schools if it means those schools could leapfrog them on league tables. A lower ranking for a school could result in a decline in enrolments, less financial resources as students leave and greater difficulty in holding and attracting staff.

There is also evidence from overseas that focus on school results undermines co-operation and collaboration between teachers within schools. The pressure to improve school results

can generate a climate of fear amongst teachers, undermine trust and result in a breakdown of the professional and social relationships needed to sustain collective professional support for student development.

Increases social segregation and inequity in education

Publication of school results and league tables tends to increase social segregation in and inequity in schooling. This is driven by both schools and parents.

Publishing school results and league tables create greater incentives for some schools to choose their students to maximise their results. There is abundant evidence from overseas of schools of “cream skimming” students most likely to achieve good results – these students are generally from higher income families.

Many parents see education as a “positional good” in that the value of education depends not on the learning acquired but on the relative status of the school attended. Publishing tables of school results aids this search for status and self-segregation. In general, it is well-off families who make greatest use of choice of school and finding the "best" schools inevitably means those that have less low-income and minority students.

Studies show that students from higher income families are far more likely to transfer to wealthier school districts and that white students are more likely to opt out of racially diverse schools and transfer to those with greater percentages of white students.

Increased social segregation in schooling induced by choice and competition between schools and aided by publication of school results and league tables can exacerbate inequity in education in two main ways.

First, it increases disparities between schools in student learning needs and the real resources available to meet those needs. Low SES schools are generally funded on the same per capita basis as other schools, with few allowances for the level of need they have to deal with. They have less real resources because they have higher costs and burdens. They also tend to have less experienced, less qualified teachers.

Second, increasing concentrations of students from low socio-economic status (SES) families in some schools tend to lead to lower overall outcomes. Schools with high concentrations of socio-economically and educationally disadvantaged students often have detrimental effects on student achievement. A student attending such a school is likely to have lower outcomes than a student from a similar background attending a school where the average SES of the student body is high.

This impact also appears to be greater for low SES and immigrant and minority students. There is a "double jeopardy" effect for these students in that they tend to be disadvantaged because of their circumstances at home, but when they are also segregated into low SES and/or predominantly minority schools this disadvantage is compounded.

A further effect of increased segregation in schools by class, religion and race is to make it more difficult for children to develop a real understanding of people of different backgrounds and to break down barriers of social intolerance. Increased social segregation means that more and more students have less experience of mixing with and learning with children from different socio-economic, cultural and ethnic backgrounds.

1. Introduction

The weight of evidence from the major research studies over the past 20 years is that publication of school results and league tables do not increase student achievement. But, far from improving school results, there is also strong evidence that they do significant harm to education. Overseas experience shows that they:

- Narrow the curriculum;
- Distort teaching practice;
- Disadvantage low and high achieving students and unfairly stigmatise low achieving students;
- Make it more difficult for low performing schools to retain high quality teachers;
- Discourage co-operation and collaboration between schools and teachers; and
- Increase social segregation and inequity.

There is some evidence of this already happening in Australia after only one year of publishing school results.

2. Narrowing the curriculum

Publishing school results creates incentives for schools to narrow the curriculum and provide a less rounded education. There are three major aspects of narrowing the curriculum:

- More time and resources is devoted to the tested subjects while other subjects not tested are neglected;
- In the subjects that are tested, greater emphasis is given to the areas that are most conducive to testing by multiple-choice questions and there is less teaching of more complex thinking and writing skills

Overseas evidence shows that schools direct more time into the tested subjects of literacy and maths while untested subjects such as science, history, social studies, languages, arts and music, physical education and health receive much less time. Even recess gets cut.

A major review of the primary school curriculum in England found that testing and publishing school results in English and maths has distorted children's learning and eroded their entitlement to a broad education [Alexander 2009; see also Alexander 2010]. It said that children were receiving an education that was "fundamentally deficient". It was neither broad nor balanced, and it valued memorization and recall over understanding and inquiry. The review found upper primary school students spend around half their time in the classroom studying English and maths because of the pressure to pass national tests. As a result, other subjects such as history, geography, art and science have been "squeezed out" of the curriculum.

> ....as children move through the primary phase, their statutory entitlement to a broad and balanced education is increasingly but needlessly compromised by a 'standards' agenda which combines high-stakes testing and the national strategies' exclusive focus on literacy and numeracy.
> The most conspicuous casualties are the arts, the humanities and the kinds of learning in all subjects which require time for talking, problem-solving and the extended exploration of ideas. [CPR: 22]

The leader of the review, Professor Robin Alexander, has said that the narrowing of the curriculum may have even reduced overall standards:

> The Cambridge Review's evidence shows how the pursuit of this narrow concept of 'standards' at the primary stage, in which test scores in literacy have been treated as proxies for the quality of primary

> education as a whole, has over the past 13 years seriously compromised children's legal entitlement to a broad and balanced curriculum. We also consider it possible that because standards in the basics and the availability of a broad and balanced curriculum have been shown empirically to be linked, the narrowing of the curriculum in pursuit of standards in 'the basics' may have had the opposite result to that intended, *depressing* standards in 'the basics' rather than raising them. As collateral damage goes, that's pretty spectacular. [Alexander 2010: 6]

The study confirmed the findings of other academic studies. For example, an earlier study of longitudinal data on the national curriculum from 1997 to 2004 found a primary curriculum skewed in the direction of English and mathematics to the detriment of science, the humanities and the arts [Boyle & Bragg 2006]. English and mathematics had increased in percentage teaching time while science, geography and history had decreased. It said this was caused by a range of central policy requirements including national testing and publishing its results.

The UK Qualifications and Curriculum Agency (QCA) submitted to a House of Commons inquiry on testing and assessment that 90% of primary and 79% of secondary schools had reported to the QCA that national testing has led to students being offered a narrower curriculum [HC 2008b: Ev 22]. The former head of the QCA has said:

> In England, the government's use of the key stage tests has seriously damaged the breadth and quality of primary education....As a result the school curriculum is narrower and poorer than it was when the tests were introduced in 1997. In many schools, the time spent on areas of the curriculum which are not externally assessed has contracted sharply. [Boston 2009: 4-5]

The Commons inquiry report concluded:

> A creative, linked curriculum which addresses the interests, needs and talents of all pupils is the casualty of the narrow focus of teaching which we have identified. Narrowing of the curriculum is problematic in two ways: core subjects are emphasised to the detriment of other, important elements of the broader curriculum; and, for those subjects which are tested in public examinations, the scope and creativity of what is taught is compromised by a focus on the requirements of the test. We are concerned that any efforts the Government makes to introduce more breadth into the school curriculum are likely to be undermined by the enduring imperative for schools, created by the accountability measures, to ensure that their pupils perform well in national tests. [HC 2008a: para 140]

There is similar evidence of an increasingly unbalanced curriculum in the United States as a result of the pressure to lift test results in English and mathematics.

A study prepared for the U. S. Department of Education as part of its requirement under the No Child Left Behind legislation to report to Congress on the effects of the law found substantial evidence that time for non-tested subjects, such as art, social studies and foreign languages was sacrificed in favour of tested subjects [Hannaway & Hamilton 2008].

> Not surprisingly, the available evidence confirms that, under conditions of a performance-based accountability system, instructional time on tested subjects increases. It also appears that this increase sometimes comes at the expense of non-tested subjects, resulting in a narrowing of the curriculum. [7]

A report by the National Centre for Education Statistics found that instructional time in grades 1 through to 4 on English and mathematics increased between 1987-88 and 2004-04 but declined in science and social science [Morton & Dalton 2007].

A study by the US Centre on Education Policy has showed that since the No Child Left Behind Act was enacted in 2002 to require more testing and reporting for reading and mathematics, by 2007 58% of all school districts had increased the time spent in primary

schools on English and 45% had increased the time spent on mathematics [McMurrer 2008]. The average increase across these districts amounted to 141 extra minutes per week (or an average of 28 minutes per day) in English and 89 extra minutes per week (or about 18 minutes per day) in maths. These amounted to increases of 47 and 37% respectively.

The time spent on social studies was reduced in 36% of school districts while time on science was reduced in 28% of districts and 20% of districts cut recess time. Time on social studies, science, art and music, gym and recess was cut by an average of 145 minutes a week. This amounted to about a one-third reduction in time spent on these subjects.

A RAND study based on surveys in California, Georgia and Pennsylvania found most teachers reporting a narrowing of the curriculum in response to state testing requirements [Hamilton et.al. 2007]. Teachers reported increased time devoted to English and mathematics, most often at the expense of time spent on social studies, arts and music and physical education. A series of case studies of schools in Illinois, Rhode Island and Washington also found that schools had increased time spent on English and mathematics in response to federal and state testing and reporting requirements [Srikantaiah 2009].

Some detailed analytical studies have also found evidence of moving resources from subjects not tested to the subjects that were tested in response to accountability requirements. For example, one study compared trends in mathematics and reading achievement after the introduction of high-stakes testing in Chicago with test score trends in social studies and science, subjects that are not included in the Chicago accountability policy [Jacob 2005]. The gains in mathematics and reading were roughly two to four times larger than gains in science and social studies leading the author to conclude that schools may have shifted resources across subjects.

A recent study of the impact of high-stakes testing in Texas found evidence of strategic resource shifting across subjects within classrooms from those not tested to those that are tested [Reback 2008]. It also found that schools were improving student performance in tested subjects at the expense of performance in other subjects.

Apart from shifting resources between subjects, there is also evidence that schools respond to testing and reporting requirements by more time spent on basic skills teaching and less teaching of more complex skills. Of major concern is that the tests used to measure school performance largely resort to multiple-choice questions and schools increasingly constrict what is taught to skills that are most conducive to testing by multiple-choice.

This approach reduces both the breadth and depth of subject learning. There is a reduction in time spent on exploring more imaginative and creative aspects of the curriculum which are not easily reducible to multiple-choice questions. Shallow learning is also emphasised at the expense of deeper learning and understanding. There is an emphasis on short-term memorisation and 'test tactics' rather than deep learning and understanding

Concerns about the narrowing of subject areas in response to state testing and publication of school results have been expressed in several US studies. For example, a report to the US Education Department noted that there appears to be a widespread tendency for schools to adopt instructional materials and practices that mirror the format of state tests [Hannaway & Hamilton 2008].

In summary, then, there is considerable evidence that the combination of standardised testing and publication of school results leads to a less well-rounded education. The House of Commons inquiry on testing and accountability took considerable evidence on the narrowing of the curriculum and its report concluded:

> ....the way that many teachers have responded to the Government's approach to accountability has meant that test results are pursued at the expense of a rounded education for children. [HC, para 130]

Diane Ravitch, Professor of Education at New York University, who has recently reviewed the experience with testing and accountability programs in the United States concluded:

> When schools are incentivized to measure only basic skills, then everything else loses time and is de-emphasized: the arts, history, geography, civics, science, foreign languages, even physical education. When the test results are used to reward or punish teachers, principals, and schools, then there is even less time for anything that is not tested. When education becomes warped in this way, quality goes down. [Ravitch 2010]

There is already some evidence that publishing NAPLAN school results on My School is narrowing the curriculum in secondary schools in Australia. A survey conducted by the Australian Secondary Principals Association found that 33% of principals said that publication of school results had reduced the breadth of curriculum in their school [ASPA 2010]. Several comments were submitted in the survey about the impact on curriculum. For example:

> …. the effect of the NAPLAN comparisons will be negative. Staff spend more time teaching to the test and less on higher-level thinking skills. The curriculum is narrowed and a school's worth is limited to the average of these 1 hour test scores.
> I have taken periods out of the curriculum in years 7 and 9 to teach to the test. The destructive impact of the simplistic colour coded results of "school effectiveness" is too damaging and has forced this regression to "Educational Fundamentalism" which in itself is shameful. [ASPA 2010: 10]

The survey results indicate that the narrowing of the curriculum is occurring within subject areas. There is increasing emphasis on literacy and numeracy within existing subject offerings and some principals reported that they have reduced the number of subjects offered to students as a result of publication of school test results on the My School website.

3. Distorting teaching practices

Teaching practice also tends to change under the pressure of publishing school results and the pressure to improve league table rankings. Schools and teachers tend to focus more on teaching test-taking skills and practicing for tests at the expense of deeper learning experiences that develop analytical skills and greater understanding.

A widespread phenomenon is “teaching to the test”. This is a term subject to much misunderstanding. It is sometimes observed that so-called “teaching to the test” is not a problem if the test reflects the breadth and depth of the curriculum. However, the problem is that it is difficult to adequately reflect the breadth and depth of the curriculum in a multiple-choice test format of about 30-40 questions to be completed over 45-60 minutes. Schools and teachers emphasise teaching in those areas most conducive to multiple-choice questions and focus on imparting test taking skills and practising for tests.

“Teaching to the test” is better understood as the approach to teaching which emphasises the learning of skills that increase test scores without increasing the underlying skills and knowledge that the test was designed to measure. There are at least two aspects of teaching to

the test. One is drilling of students in the likely questions that they will encounter in the tests through test practice. The other is instructing students in test-taking skills.

In teaching to the test a high proportion of teaching time is given over to test preparation. Teachers coach students on test technique, question spotting, going over sample questions similar to those likely to be set in the test and generally focussing teaching of the substance of a subject in a way best calculated to maximise marks in the test.

There is considerable anecdotal evidence about teaching to the test. For example, a large majority of those who presented evidence to the House of Commons inquiry on testing and accountability were clear that teaching to the test is prevalent and that it is caused by the serious consequences for schools attached to the publication of the results of the tests [HC para 116]. The report concluded that the phenomenon is "widespread" [HC para 130].

Last year, the director of the Oxford University Centre for Educational Assessment and former head of the NSW Board of Studies, Gordon Stanley, warned Australian education authorities about the dangers of teaching to the test. He said:

> We could well end up with a similar situation to the UK, where you get a whole industry created around improving performance on the tests rather than necessarily improving students' learning skills....That has led to a lot of teaching to the test and schools focusing on kids who are close to achieving the targets on the view that they are going to be the easiest to improve. [cited in Patty 2009]

National and state surveys of teachers in the US indicate that teachers in states with high stakes accountability regimes spent a considerable amount on test preparation in class. For example, one survey found that teachers in states with high-stakes tests at the time were much more apt than their counterparts in states with lower-stakes tests to engage in test preparation earlier in the school year; spend more time on such initiatives; target special groups of students for more intense preparation; use materials that closely resemble the test; use commercially or state-developed test-specific preparation material; and use items from previous tests for practice by students [Pedulla et.al. 2003].

Moreover, this emphasis on improving test taking skills is often reinforced by principals and school administrators. For example, a recent study found that about 90% of principals in Georgia and Pennsylvania distributed test preparation materials such as practice tests to teachers, and similar percentages distributed released copies of the state tests [Hamilton et. al. 2007]. Almost all principals said they addressed test preparation in staff meetings and helped teachers identify content that is likely to be on the state test.

This focus on developing test taking skills has reached extreme proportions in Washington DC. In 2008, schools chancellor, Michelle Rhee launched a new program to improve student achievement in Washington DC called the Saturday Scholars program. Under this program, selected students are invited to come to school on Saturdays to work on their reading, math and test preparation skills. In announcing the new program Rhee said its main goal is to "instill lifelong testing competence" [Rhee 2008].

Under this regime, improved education is not about developing lifelong learning, but lifelong testing skills. A broad education and deep learning is not the priority.
Publishing school results and league tables turn classrooms into test preparation factories. In schools all over England and the United States, weeks and months are devoted to test preparation instead of deeper learning. For example, the former head of the UK

Qualifications and Curriculum Agency, Ken Boston, recently commented on the situation in England as follows:

> The amount of time spent on test preparation has increased over the past 10 years: in the second half of the spring term 70 per cent of schools spent more than three hours per week on test preparation. In some extreme cases, months have been spent in the final year of primary schooling on nothing else than test preparation, to the neglect of the other areas of the curriculum and hence to the great detriment of the quality of the children's education. [Boston 2009: 5]

There is considerable evidence that publication of NAPLAN school results on the My School website and publication of school league tables in the media is causing schools to spend more time on test practice in class.

A survey conducted by the Australian Secondary Principals Association found that 65% of schools reported that they had increased the time spent in class on preparation for the NAPLAN tests in 2010 and 70% said they had increased the time spent on practising tests [ASPA 2010].

A report published by the Australian Primary Principals Association said large amounts of valuable instructional time was taken up by coaching and practising tests [APPA 2010]. Some schools spent a significant amount of time on test preparation techniques and allocating test practice as homework to improve their NAPLAN results in 2010. This consumed valuable time that should have been spent on literacy and numeracy and other important areas of the primary curriculum.

There were several reports following the NAPLAN tests in 2010 of schools spending disproportionate amounts of time on test practice to improve school results to the detriment of other areas of the curriculum. The President of the Australian Secondary Principals Association, Sheree Vertigan, said that most students were spending "a lot of time" in class preparing for the NAPLAN tests and that it was "defeating the whole purpose of the tests in the first place" [*The Australian* 10 May 2010]. She said encouraging students to study for NAPLAN tests was "manipulating" the data and was the main catalyst for "narrowing the curriculum".

The Queensland Teachers' Union president Steve Ryan said schools were spending too much time grooming students in an attempt to make them perform better in the NAPLAN tests.

> We've got the ridiculous situation of schools just setting aside a whole range of good curriculum offerings just to concentrate on NAPLAN so the school would be seen to be in some way better than the school next to it. [*ABC News*, 11 May 2010]

The principal of one Perth private school said that it was a "sad week for education":

> The NAPLAN tests are now almost completely useless as a diagnostic tool for individual student performance as different schools choose to prepare so differently, many coaching to the test to avoid potential public humiliation...Stories already abound of a narrowed curriculum filled with NAPLAN lessons from day one, term one from Year 2 onward. [*West Australian* 12 May 2010]

Some schools had prepared for the tests for longer than six months. Numerous teachers told *The Courier-Mail* they spent more than half of their class time on exam preparation and practice questions for NAPLAN [12 May 2010]. The *Canberra Times* reported that some ACT schools had spent up to 90 minutes a day over two weeks practicing for NAPLAN [11

May 2010]. Several parents at a south Canberra school said that students had to sit for two practice tests a day in the lead-up to the NAPLAN tests.

There were also several similar reports about the 2009 NAPLAN tests, the results of which were reported on the My School website in 2010. For example, a report by the Queensland Studies Authority analysing the 2009 test results warned teachers about over-practising for the writing exam. *The Courier-Mail* reported that test markers felt that Queensland students had "over-practised" for the 2009 NAPLAN writing task [11 May 2010]. The *West Australian* reported that up to a quarter of school time was being spent on preparing for the tests and that other subjects such as science had been temporarily dropped to allow more time to practice tests [14 April 2009].

It is reported that many teachers refer to NAPLAN as “napalm” because “it kills everything in the classroom” [*The Courier-Mail*, 18 March 2009].

There is also evidence that school administrators in some states have pressured principals and schools to devote more time to practising for tests. For example, the Victorian Education Department has taken an active role in pressuring principals and teachers to practice for NAPLAN. A memo sent to schools in rural Victoria at the beginning of the 2010 school year told teachers to "explicitly teach" for the national literacy and numeracy tests as part of a drive to lift the state's overall performance [*The Australian* 5 February 2010]. It outlined a 10-week strategy for the department and schools and a "delivery strategy" for teachers in the lead-up to the tests in May. Principals were told to appoint a NAPLAN co-ordinator, set up a sample testing process that may "require resourcing" and "provide additional assistance for students identified as capable of making significant improvement". A "blueprint for classroom approaches" set out strategies to coach students in skills for passing tests, including learning the "test question vocabulary", "skim and scan", and to skip questions that confuse them.

In April last year, the head of the Victorian Department of Education, Peter Dawkins, sent a memo to all principals suggesting more time be spent on preparing students for NAPLAN so as to improve Victoria’s results. The memo stated:

> Students need assistance with preparation to effectively participate in the NAPLAN assessment. They have to understand the genre of testing and the cognitive demands they will be placed under to successfully complete the task. [Dawkins 2009]

The Australian Primary Principals Association president Leonie Trimper said that principals in South Australia were giving consideration to intensive training for students for NAPLAN because they had been placed under extreme pressure from district managers to improve their test results [*The Advertiser*, 30 March 2010].

The president of the Queensland Association of State School Principals, Norm Hartt, said that education officials had told principals to lift their results “at all costs” [*Courier-Mail* 10 June 2010]. Last year, it was also reported that Queensland education officials were putting tremendous pressure on teachers to lift results by practising for tests [*Courier-Mail* 18 March 2009]. Schools were told to put extra time into practising for the tests, even if it was at the expense of time spent on other subjects. The Queensland Teachers Union said that teachers were told that if the results don’t improve, their own employment positions would be reviewed.

The NAPLAN website itself encourages practising for tests. It says that test preparation increases student understanding of what is being asked and provides strategies to focus on test content.

Teaching to the test has even been condemned by Julia Gillard's own deputy chief of staff, Tom Bentley, who was an advisor to the first Education Secretary of the Blair Government:

> The premium placed on test results has encouraged schools and teachers to teach "to the test". This reinforces a system in which students are offered few real incentives to transfer skills across disciplines and contexts or solve real problems within disciplines - to develop their understanding in ways which they could apply in the world beyond the exam hall. [Bentley 2002]

Yet, this is precisely the incentive that his Minister has created for teachers and principals in schools across Australia. It will only do harm to Australian students by restricting and narrowing their learning.

A longer term effect of a focus on test preparation and raising test scores is that it comes to influence ideas about "good teaching". Some studies show that the idea of a "good teacher" changes in schools under the pressure to lift school results. A "good teacher" becomes someone who raises test scores and attains a high pass rate [for example, see Booher-Jennings 2005]. The idea of what it is to be a "good teacher" is itself distorted and corrupted by the pressure to improve test scores.

This development contradicts and endangers the effort to improve teacher quality through recognised standards and professional development. The likelihood is that the ever-present focus on raising test scores will predominate over other notions of "good teaching", and thereby undermine efforts to improve the quality of teaching.

4. Disadvantages low and high achieving students

The pressure to improve school results can also create incentives for schools to ignore students at the lower and upper ends of the student achievement scales. Several English and US studies have found that schools concentrate on improving the results of those students who are on the border of accepted benchmarks and neglect the lowest and highest achievers. This practice is often referred to as "educational triage".

Improving the results of students just below benchmarks is seen as the most efficient way to improve a school's average score or the proportion of students achieving a benchmark. However, while this strategy gives schools a better chance of improving their ranking, it may lead to worse outcomes for low and high performing students.

A couple of anecdotes illustrate how "educational triage" is implemented. According to a *Washington Post* report, the principal of a Washington DC school handed out a list of all the black, Hispanic, special-education and limited-English-speaking students who would take the Maryland School Assessment, the measure of success or failure under the federal No Child Left Behind (NCLB) legislation.

> We were told to cross off the kids who would never pass. We were told to cross off the kids who, if we handed them the test tomorrow, they would pass. And then the kids who were left over, those were the kids we were supposed to focus on. [De Vise 2007]

Those who remained on the list received 45 minutes of intensive test preparation four days a week prior to the tests.

Another anecdote comes from a Texas study where a consultant told teachers:

> Take out your classes' latest benchmark scores …. and divide your students into three groups. Color the 'safe cases,' or kids who will definitely pass, green. Now, here's the most important part: identify the kids who are 'suitable cases for treatment.' Those are the ones who can pass with a little extra help. Color them yellow. Then, color the kids who have no chance of passing this year and the kids that don't count—the 'hopeless cases'—red. You should focus your attention on the yellow kids, the bubble kids. They'll give you the biggest return on your investment. [Booher-Jennings 2006: 757]

Such anecdotal evidence is confirmed by many studies. Booher-Jennings [2005] provides a rich case study of this practice in a Texas school. A recent RAND study of how educators responded to the NCLB accountability requirements in California, Georgia, and Pennsylvania found that some teachers reported focusing more on students near the proficiency cut-off score and expressed concerns about the accountability system's negative effects on the curriculum and instruction provided to high-achieving students [Hamilton et.al. 2007; see also Hamilton & Berends 2006].

A recent study of the impact of the NCLB in the state of Washington found that it provided incentives for schools to focus educational resources on the marginal student rather than those on the tails of the achievement distribution [Krieg 2008]. It found that in schools more likely to be sanctioned under the NCLB requirements, students at either end of the distribution tails make fewer gains than would be expected if their school did not face sanctions. Further, students in the middle of the achievement distribution do better than expected on the high-stakes exam if they attend schools that are likely to be sanctioned.

A study which analysed test scores from 5th grade students in the Chicago Public Schools found that after NCLB was implemented, students performing near the middle of the score distribution showed larger gains than students performing above that level. The gains among the lowest-scoring students were mixed and not as consistent as those of the students scoring near the middle [Neal & Schanzenbach 2010].

Similar results have been found in England. A study using secondary school data found that schools focus on improving the performance of students just below pass benchmarks and that this policy reduced the educational gains and exam performance of very low achieving students [Burgess et.al. 2005].

However, there is some contrasting evidence. Ballou and Springer [2009] found evidence of gains to students at the bottom of the distribution, but find no consistent evidence that schools facing accountability pressure neglect their high achieving students to focus on low achievers. A study using Texas data from the 1990s found that, contrary to the triage hypothesis, when a school has a realistic chance of improving its accountability rating, the lowest performing students make greater than expected gains, even if they have no chance of passing the exam in that subject [Reback 2008]. Nevertheless, it did find some evidence of adverse effects on high achieving students. Another study using North Carolina reading and mathematics test data found little evidence that schools with low results ignore very low achieving students in their efforts to improve their overall school results [Ladd & Lauen 2010]. However, there was some evidence that reading results were reduced for the higher achieving students.

There is also some evidence of in Australia that schools have responded to the higher stakes associated with the NAPLAN test by concentrating on those students most likely to show

improvement if given extra assistance [APPA 2010]. Schools allocated more resources to this select group of students and other students with greater needs did not receive as much attention for the first five months of the year until after completion of the NAPLAN tests.

5. Unfairly stigmatises students and schools

Publication of the results of individual schools and public rankings of schools often unfairly stigmatise and humiliate schools, teachers, students and their families.

The *My School* website uses colour codes to grade schools on their results in four literacy domains and numeracy for a school in each Year level tested. Results substantially below average are to be graded 'red'; those below average are to be graded 'pink'; schools around average will be graded 'white'; those above average will be graded 'pale green' and those substantially above average graded 'green'.

The Federal Education Minister gave an absolute guarantee that *My School* would not "name and shame schools". Yet, this is precisely what it does. Schools with the lowest results are to be flagged 'red', meaning 'red for danger'. It is a signal to parents to keep away. This is public shaming of the worst possible kind.

By 'red flagging' schools, the Government is aiding and abetting an annual ritual hunt for the worst performing schools around Australia that happens in England and the United States. It will unfairly condemn schools in the most difficult circumstances and make their task harder.

Many schools serving the most disadvantaged communities in Australia have been given red flags on My School. They are the schools working in the most difficult of circumstances, yet their reward from the Rudd Government is to be pilloried in public.

Giving schools a 'red flag' is designed to punish. Why else was the colour red chosen for these schools? Punishing schools by publicly labelling them as 'failures' in this way is not the path to school improvement. It is likely to be counter-productive by undermining teaching and learning.

National tests are conducted at two Year levels only in both primary and secondary schools. It means that students in these cohorts are easily identifiable as the "culprits" for a school receiving a 'red flag', especially in small schools, of which there are many in rural areas of Australia. These students will be humiliated and demoralised.

This is not the way to encourage improvement. Students who are humiliated for their learning accomplishments are unlikely to respond positively in their future learning. They could become alienated from schooling and, possibly, from their peers and their community as well.

The task of teachers and schools that are 'red flagged' is made much harder. Far from creating incentives for better performance, 'red flagging' schools is more likely to impair the future learning of some students.

Labelling schools as failures is also likely to set off a spiral of decline. It may cause some parents to leave the school. Reduced enrolments make it harder to retain adequate resources and to keep and recruit good teachers. It could well lead to lower student achievement and lower average school results.

6. Harder for low performing schools to retain teachers

A common perception is that publishing school results, school league tables and other accountability measures make it more difficult for schools with low results to retain and attract quality teachers. While many studies have examined the difficulty high poverty schools have in retaining and attracting quality teachers, only a few have considered the impact of accountability measures in this regard. The weight of evidence in these studies suggests that low performing schools have greater difficulty in retaining teachers.

One study investigated the impact of the introduction of North Carolina's school accountability program [Clotfelter et.al. 2004]. It found that teachers in low-performing schools left teaching at a higher rate in the period following the implementation of the program than in the period before. However, the study did not find evidence that teacher quality was affected by the program. The percentage of quality teachers remained higher in low-performing schools than in middle-to-high-performing schools both before and after the implementation of the program, and no statistically significant differences emerged in the changes that followed implementation.

A review of empirical studies of teacher recruitment and retention made a tentative finding that accountability policies might lead to increased teacher attrition in low-performing schools [Guarino et.al. 2006]. A recent study of the relationship between school accountability and teacher mobility in Florida found strong evidence that teachers are more likely to leave schools that have received a low grading and they are less likely to leave schools that received a high grading [Feng et.al. 2010].

A contrasting result was found in a study of 4th grade teachers in New York State public elementary schools from 1994-1995 through 2001-2002 which examined the response of teachers to the implementation of state-mandated testing. It found that teachers were no more likely to leave schools with low results [Boyd et.al. 2008].

A unique study using data from North Carolina schools examined whether accountability causes reallocation of high-performing principals across schools [Li 2010]. It found that after NCLB was introduced high ability principals at schools more likely to face sanctions for inadequate improvement in student achievement disproportionately move to schools less likely to face sanctions. These changes in the assignment of schools to principals translate into substantive declines in principal effectiveness at schools serving disadvantaged student populations.

7. Discourages co-operation and collaboration

Collaboration is generally seen as an important way to spread innovative approaches and good teaching practice both between and within schools [for example, see Ainscow & West 2006]. Publishing school results and competition for rankings can reduce collaboration between schools and teachers and slow the dissemination of best practice.

First, schools will be reluctant to share successful practices with other schools if it means those schools could leapfrog them in ranking on league tables. For example, a survey by the New Zealand Council for Educational Research found that 38 per cent of schools who faced competition were prepared to share resources or offer mutual support to other schools compared with 58 per cent of schools who did not face competition [Wylie 2006]. A lower ranking for a school could result in a decline in enrolments, less financial resources and greater difficulty in holding and attracting staff.

Second, there is also evidence from overseas that focus on school results undermines co-operation and collaboration between teachers within schools. For example, one detailed Texas case study noted the practice of principals presenting faculty meetings with a "league table" ranking teachers by the percentage of their students passing the latest benchmark test. This engendered a competitive environment within schools and a climate of fear at being labelled a "bad" teacher [Booher-Jennings 2005; 2006]. This put a serious strain on teacher relationships within the school and undermined trust between teachers. Putting teachers into competition with one another led teachers to see their colleagues as threats rather than partners. It served to individualize teacher performance rather than encouraging collective responsibility for all students.

8. Increases social segregation and inequity in education

Publishing school results and publishing league tables tends to increase social segregation and inequity in education. This is driven by both schools and parents. Not only does it increase social division, but it may also harm overall education outcomes, especially for students from low socio-economic status families.

Publishing school results and league tables create greater incentives for some schools to choose their students to maximise their results. Many schools actively choose their enrolments by "cream skimming" students most likely to achieve good results – these students are generally from white, well-off families. Selective and high demand government schools and private schools have control over their enrolments and this allows them greater opportunities to select higher achieving students.

There is abundant evidence from overseas of schools selecting their students to maintain or improve their league table position. For example, a favourite strategy is to use formal and informal enrolment criteria to "cream skim" or "cherry pick" students most likely to achieve good results. In England, the misuse of school admission procedures to maximize school results has been a major ongoing issue since the introduction of league tables. Various methods have been used to select certain groups of students and exclude others, such as giving priority to the children of employees, former students, those with a family connection to the school, and selecting a proportion of children on the basis of aptitude/ability in a subject area(s) or on the basis of general ability [West et.al. 2004]. Many schools also required parents to provide supplementary information unrelated to the school's admissions criteria, such as their occupation, whether the family lived in a hostel or bed and breakfast accommodation and whether parents had refugee status. Many of these practices continue despite government efforts to stamp them out [West et.al. 2009].

A recently published study of public sector religious secondary schools in London has found that selective 'élite' schools appear to 'select out' low income religious families, thereby displacing them to religious schools with a less affluent composition [Allen & West 2009]. It identified a range of different admissions criteria and practices used by the socially selective schools including school-administered banding, aptitude tests, tests of religious or denominational commitment, primary school references and others that may have contributed to the under-representation of lower ability pupils. Moreover, the hierarchy of schools that has developed also tends to dissuade many parents from even applying to the more selective schools because they believe they will not meet the selection criteria.

> ...schools' admissions criteria and practices are important determinants of which pupils are offered places, whether on account of schools attracting applications from certain parents or in terms of the admissions process itself. [Allen & West 2009: 19]

This experience has direct relevance for Australia where private schools have control over their enrolments. Publishing school results and league tables are likely to lead to even greater selection to protect school reputations and rankings. Such selection processes may also extend to the government sector where some schools are in such high demand that they can effectively discriminate amongst applicants for enrolment. Australian governments are unlikely to even attempt to stamp out enrolment discrimination by private and government schools.

On the other hand, many parents see education as a "positional good" in that the value of education depends not on the learning acquired but on the relative status of the school attended. Publishing tables of school results aids this search for status and self-segregation.

In general, it is well-off, white families who make greatest use of choice of school. Extensive evidence from several countries shows that the parents who make use of school performance comparisons and actively choose schools are better educated and have higher levels of income than those who do not [for example, see Cullen et.al. 2005; Hamilton & Guin 2005]. Publishing school results serves these higher income families.

For high income families, finding the "best" schools inevitably means those that have the fewest low-income and minority students. This is demonstrated by a wide range of studies for several cities and regions in the United States and in other countries such as England, Netherlands and New Zealand [see Cobbold 2007]. For example, studies published by the US National Bureau of Economic Research have found that public school choice tends to increase education stratification [Hastings et.al. 2005, 2006, 2009].

A recent study of school choice amongst elementary and middle school students in Durham, North Carolina, compared the characteristics of students who opt out of their assigned school with students who do not [Bifulco et.al. 2009]. It found that well-off families tend to opt out of assigned schools with concentrations of disadvantaged students and to attend schools with higher achieving students. Another study of school choice in Denver, Colorado, found that it was higher income families who were far more likely to participate in inter-district choice and were much more likely to transfer to wealthier school districts. White students were more likely to opt out of racially diverse districts and transfer to districts with greater percentages of white students [Holme & Richards 2009]. Other studies have found that white families tend to choose schools on the basis of their racial composition [Bifulco & Ladd 2007; Saporito 2009]. Another found that parents' decision strategies appeared to be associated with social class status [Bell 2009].

Thus, parents who exercise school choice tend to choose schools with peer compositions that are similar to their own race and social status. By aiding parent choice, publication of school results and league tables can reinforce tendencies toward racial and socio-economic separation in schools.

Increased social segregation in schooling can exacerbate inequity in education in two main ways.

First, it increases disparities between schools in student learning needs and the real resources available to meet those needs. Schools with a high proportion of students from low socio-economic (SES) families have higher levels of learning needs and other problems than high SES schools because low income is associated with lower levels of student achievement. In low SES schools, the scale of challenges is much larger because of the greater concentration of students experiencing them is greater. They generally have to devote far more time and resources to family and health crises, children with few educational materials in their homes, and many children with very weak educational preparation.

Generally, the resources available to low SES schools are not commensurate with the problems they face. Low SES schools are generally funded on the same per capita basis as other schools, with few allowances for the level of need they have to deal with. They have less real resources because they have higher costs and burdens.

Second, increasing concentrations of students from low SES and minority families in some schools tend to lead to lower overall outcomes. Many international studies show that there is a school composition effect on student outcomes associated with high proportions of students from low SES and minority families [see Borman & Dowling 2010; Dronkers & Levels 2007; Oh 2007; Rangvid 2007; Rumberger & Palady 2005, 2006]. A student attending a school where the average SES of the student body is low is likely to have lower outcomes than a student from a similar background attending a school where the average SES of the student body is high. Thus, increasing social segregation between schools is likely to reduce average student achievement.

Furthermore, while the social composition of schools has a significant impact on the achievement of all students, studies show that the impact is greater on low SES and immigrant or minority students [Willms 2006]. There is a "double jeopardy" effect for students from low SES and minority families in that they tend to be disadvantaged because of their circumstances at home, but when they are also segregated into low SES and/or predominantly minority schools they are likely to fare even worse.

Increasing social segregation in schools also has a further effect on the education of the young in society. Schools segregated by class, religion and race make it more difficult for children to develop a real understanding of people of different backgrounds and to break down barriers of social intolerance. Socially segregated schools can feed a lack of social empathy, indeed, social intolerance and an inability of people from different backgrounds to learn to work together and live in the same communities together.

References

Ainscow, M. and West, M. (eds.) 2006. *Improving Urban Schools: Leadership and Collaboration*, Open University Press, Buckingham.

Allen, R. & West, A. 2009. Religious Schools in London: School Admissions, Religious Composition and Selectivity. *Oxford Review of Education*, 35 (4): 471-494.

Alexander, R. (ed.) 2009. *Children, Their World, Their Education*. Routledge, London.

Alexander, R. 2010. The Perils of Policy: Success, Amnesia and Collateral Damage in Systemic Educational Reform. Alexander Miegunyah Lecture, Melbourne Graduate School of Education, University of Melbourne, Melbourne, 10 March. Available at: http://www.education.unimelb.edu.au/news/news%20pdfs/Alexander%20Miegunyah%20lecture%20FINAL.pdf

Australian Primary Principals Association (APPA) 2010. The Reporting and Use of NAPLAN, June. Available at: http://www.appa.asn.au/index.php/appa-business/news-items/993-appa-wants-naplan-reform-media-release

Australian Secondary Principals Association (ASPA) 2010. My School Survey March 2010 Results. Available at: http://www.aspa.asn.au/index.php?option=com_content&view=article&id=369:report-from-the-my-school-impact-on-schools-survey&catid=1:latest-news&Itemid=59

Ballou, D. & Springer, M. G. 2009. Achievement Trade-offs and No Child Left Behind. Paper presented to the *NCLB: Emerging Findings Research Conference* at the Urban Institute, Washington, DC, 12 August. Available at: http://www.caldercenter.org/nclbconference.cfm

Bell, C. 2009. All Choices Created Equal? The Role of Choice Sets in the Selection of Schools. *Peabody Journal of Education*, 84 (2): 191-208.

Bentley, T. 2002. Time to Stop 'Teaching to the Test'. *The Observer*, 10 February.

Bifulco, R., & Ladd, H. F. 2007. School Choice, Racial Segregation, and Test Score Gaps: Evidence from North Carolina's Charter School Program. *Journal of Policy Analysis and Management*, 26 (1): 31-56.

Bifulco, R., Ladd, H. F., & Ross, S. L. 2009. The Effects of Public School Choice on Those Left Behind: Evidence from Durham, North Carolina. *Peabody Journal of Education*, 84 (2): 130-149.

Booher-Jennings, J. 2005. Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, 42 (2): 231–268.

Booher-Jennings, J. 2006. Rationing Education in the Age of Accountability. *Phi Delta Kappan*, 87 (10): 756-761.

Borman, G. & Dowling, N. M. 2010. Schools and Inequality: A Multilevel Analysis of Coleman's Inequality of Educational Opportunity Data. *Teachers College Record*, 112 (5): 1201–1246.

Boston, K. 2009. A Response to the APPA Position Paper on the Publication of Nationally Comparable School Performance Data. August. Available at: http://www.appa.asn.au/images/reports/bostonaddress20090813.pdf

Boyd, D.; Lankford, H.; Loeb, S. & Wykoff, J. 2008. The Impact of Assessment and Accountability on Teacher Recruitment and Retention. *Public Finance Review*, 36 (1): 88-111.

Boyle, B. & Bragg, J. 2006. A Curriculum Without Foundation. *British Educational Research Journal*, 32 (4):569-582.

Burgess, S.; Propper, C.; Slater, H. & Wilson, D. 2005. Who Wins and Who Loses from School Accountability? The Distribution of Educational Gain in English Secondary Schools. Working Paper No. 05/128, Centre for Market and Public Organisation, University of Bristol. Available at: http://www.bris.ac.uk/cmpo/publications/papers/#a2005

Cambridge Primary Review (CPR) 2009. Introducing the Cambridge Primary Review. Available at: http://www.primaryreview.org.uk/

Clotfelter, C.T.; Ladd, H.F.; Vigdor, J.L. & Diaz R.A. 2004. Do School Accountability Programs Make it More Difficult for Low-Performing Schools to Attract and Retain High Quality Teachers? *Journal of Policy Analysis and Management*, 23(2): 251-271.

Cobbold, T. 2007. The Great School Fraud: Howard Government School Education Policy 1996-2000. Available at: http://www.saveourschools.com.

Cullen J. B.; Jacob, B. A. and Levitt, S. D. 2005. The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools. *Journal of Public Economics* 89 (5-6): 729-760.

Dawkins, P. 2009. S172-2009 National Assessment Program – Literacy and Numeracy (NAPLAN) 2009. Memo to Schools and Principals, Department of Education and Early Childhood Development – Victoria.

Dronkers, J. & Levels, M. 2007. Do School Segregation and School Resources Explain Region-of-Origin Differences in the Mathematics Achievement of Immigrant Students? *Educational Research and Evaluation*, 13 (5): 435-462.

Feng, L.; Figlio, D.N. & Sass, T. 2010. School Accountability and Teacher Mobility. Working Paper No. 16070, National Bureau of Economic Research, Cambridge, MA, June. Available at: http://papers.nber.org/papers/w16070

Guarino, C.M.; Santibañez, L. & Daley, G.A. 2006. Teacher Recruitment and Retention: A Review of the Recent Empirical Literature. *Review of Educational Research*, 76 (2): 173-208.

Hamilton, L.S.; Stecher, B.M.; Marsh, J.A.; McCombs, J.S.; Robyn, A.; Russell, J.; Naftel, S. & Barney, H. 2007. Standards-based Accountability under No Child Left Behind: Experiences of Teachers and Administrators in Three States. RAND Corporation, Santa Monica. Available at: http://www.rand.org/pubs/monographs/MG589/

Hamilton, L.S. & Berends, M. 2006. Instructional practices related to standards and assessments. Working Paper WR-374-EDU. RAND Corporation, Santa Monica, April. Available at: http://www.rand.org/pubs/working_papers/WR374/

Hamilton, L.S. & Guin, K. 2005. How Families Choose Schools. In J. R. Betts and T. Loveless (eds.), *Getting Choice Right*. Brookings Institution Press, Washington DC, 40-60.

Hannaway, J. & Hamilton, L.S. 2008. Performance-Based Accountability Policies: Implications for School and Classroom Practices. The Urban Institute, Washington DC. Available at: http://www.urban.org/publications/411779.html

Hastings, J. S.; Kane, T. J. & Staiger, D. O. 2005. Parental Preferences and School Competition: Evidence from a Public School Choice Program. Working Paper No. 11805, National Bureau of Economic Research, Cambridge, MA, November. Available at: http://www.nber.org/papers/w11805

Hastings, J. S.; Kane, T. J. & Staiger, D. O. 2006. Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery. Working Paper 12145, National Bureau of Economic Research, Cambridge, MA, April. Available at: http://www.nber.org/papers/w12145

Hastings, J. S.; Kane, T. J. & Staiger, D. O. 2009. Heterogeneous Preferences and the Efficacy of Public School Choice. Available at: http://aida.econ.yale.edu/~jh529/papers/HKS_Combined_200806.pdf

Holme, J. J., & Richards, M. P. 2009. School Choice and Stratification in a Regional Context: Examining the Role of Inter-District Choice. *Peabody Journal of Education*, 84 (2): 150-171.

House of Commons, Children, Schools and Families Committee 2008a. Testing and Assessment. Vol I: Report, London, 13 May. Available at: http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/169/169.pdf

House of Commons, Children, Schools and Families Committee 2008b. Testing and Assessment. Vol II: Oral and Written Evidence, London, 13 May. Available at: http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/169/169ii.pdf

Jacob, B. A. 2005. Accountability, Incentives and Behaviour: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics* 89 (5-6): 761-796.

Koretz, D.M. 2002. Limitations in the Use of Achievement Tests as Measures of Educators' Productivity. *Journal of Human Resources*, 37 (4): 752-777.

Krieg, J. M. 2008. Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act. *Education Finance and Policy*, 3 (2): 250-281.

Ladd, H. F. & Lauen, D. L. 2010. Status versus Growth: The Distributional Effects of School Accountability Policies. *Journal of Policy Analysis and Management*, 29 (3): 426-450.

Li, D. 2010. Unintended Consequences: No Child Left Behind and the Allocation of School Leaders. Department of Economic, Massachusetts Institute of Technology, Cambridge, MA, June. Available at: http://econ-www.mit.edu/grad/d_li/papers

McMurrer, J. 2008. Instructional Time in Elementary Schools. Centre on Education Policy, Washington DC, 20 February. Available at: http://www.cep-dc.org/

Morton, B.A. & Dalton, B. 2007. Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 Through 4. National Centre for Education Statistics, Institute of Education Sciences, Washington DC, May. Available at: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007305

Neal, D. & Schanzenbach, D. 2010. Left Behind By Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 92 (2): 263–283.

Oh, G. 2007. Effects of Socioeconomic Status of School Population on Student Math Achievement Growth. Paper presented to the Population Association of America Annual Meeting, New York, 29-31 March. Available at: http://paa2007.princeton.edu/abstractViewer.aspx?submissionId=71932.

Patty, Anna 2009. School Tests Could Fail Students: Educator. *Sydney Morning Herald*, 26 March.

Pedulla, J. J.; Abrams, L.M.; Madaus, G.F.; Russell, M.K.; Ramos, M.A. & Miao, J. 2003. *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. National Board on Education Testing and Public Policy, Chestnut Hill, MA. Available at: http://www.bc.edu/research/nbetpp/reports.html

Rangvid, B. S. 2007. School Composition Effects in Denmark: Quantile Regression Evidence from PISA 2000. *Empirical Economics*, 33: 359-388.

Ravitch, D. 2010. What I Did in June. *Bridging Differences*, 22 June. Available at: http://blogs.edweek.org/edweek/Bridging-Differences/2010/06/what_i_did_in_june.html

Reback, R. 2008. Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics*, 92(5-6): 1394-1415.

Rhee, M. 2008. Fenty, Rhee Launch "Saturday Scholars" Program to Improve Assessment Gains. Media Release, District of Columbia Government, 18 January. Available at: http://edreform.dc.gov/edreform/cwp/view,A,11,Q,460533.asp

Rumberger, R. W. & Palardy, G. J. 2005. Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High Schools. *Teachers College Record*, 107 (9): 1999-2045.

Rumberger, R. W. & Palardy, G. J. 2006. Does Resegregation Matter? The Impact of Segregation on Academic Achievement in Southern High Schools. In J. C. Boger and G.

Orfield (eds.), *School Resegregation*, University of North Carolina Press, Chapel Hill, 128-147.

Saporito, S. 2009. School Choice in Black and White: Private School Enrollment Among Racial Groups, 1990-2000. *Peabody Journal of Education*, 84 (2): 172-190.

Srikantaiah, D. 2009. How State and Federal Accountability Policies Have Influenced Curriculum and Instruction in Three States. Centre on Education Policy, Washington DC, October. Available at: http://www.cep-dc.org/

West, A.; Hind, A. & Pennell, H. 2004. School Admissions and `Selection' in Comprehensive Schools: Policy and Practice. *Oxford Review of Education*, 30 (3): 347-369.

West, A.; Barham, E. & Hind, A. 2009. Secondary School Admissions in England: Policy and Practice. Report prepared for the Research and Information on State Education Trust, London. Available at: http://www.risetrust.org.uk/admissions.html

Willms, J. D. 2006. Learning Divides: Ten Policy Questions About the Performance and Equity of Schools and Schooling Systems. UNESCO Institute of Statistics, Montreal. Available at: http://www.uis.unesco.org/ev.php?ID=6832_201&ID2=DO_TOPIC.

Wilson, D.; Croxson, B. & Atkinson, A. 2006. What Gets Measured Gets Done. *Policy Studies*, 27 (2): 153-171.

Wylie, C. 2006. What is the reality of school competition? Occasional Paper No. 126, National Center for the Study of Privatization in Education, Teachers' College, Colombia University. Available at: http://www.ncspe.org/list-papers.php.

Published School Results and League Tables are Misleading and Unreliable

Submission to the Senate Education Committee Inquiry into the Administration and Reporting of NAPLAN

Part 3

SAVE OUR SCHOOLS

June 2010

http://www.saveourschools.com.au

Table of Contents

Summary

The Federal Government claims that publishing school results on the My School website will better inform parent choice of school. However, parents can be misled by using published school results to inform their choice of school because school results are not a reliable measure of school quality.

The school results published on My School are likely to be an inaccurate and misleading measure of school quality because:

- Differences in school composition affect school results;
- Many other factors outside schools influence school results;
- They are a selective measure of education;
- They are subject to manipulation and rorting; and,
- There may be significant statistical errors on school test results.

These factors may lead parents to choose a school of lesser quality than its results indicate. They also make it difficult to identify effective school practices. Decision-makers and schools may be misled in recommending and adopting particular educational programs. Education practices and programs could be falsely identified as successes while successful programs in reality are ignored or even falsely condemned.

Differences in student composition influence school results

School results are significantly determined by the socio-economic background of school communities. School results and league table rankings are often more a measure of the family background of a school's students than the quality of its teaching.

Schools with a high proportion of students from high income families generally have higher average results than schools with a high proportion of students from low income families, but this says nothing about school quality. Test results of mediocre schools with privileged intakes can look good compared to schools with high quality teachers serving less privileged communities.

My School allows unfair and misleading comparisons between some of the most advantaged and disadvantaged schools around Australia, despite government assurances this would not happen.

Ethnic and other differences in enrolment profiles also affect school results. For example, schools with a high proportion of Chinese students generally have higher average results than those with more students from other ethnic backgrounds. Schools with higher proportions of students with disabilities participating in tests or higher proportions of Indigenous students may have lower results than other schools.

Other outside factors influence school results

School results are also strongly influenced also by other external factors such as student absenteeism, student turnover, school size, school funding, parent involvement in learning at home, and the proportion of students receiving private tutoring.

Some schools may have lower results because they have a high proportion of students who often change school. Studies show that these students tend to have lower average results than students who remain at the same school.

School results and league table rankings may significantly distorted by the results of a few students in small schools. Studies show that small schools are much more likely to report large changes in average results from one year to the next, both positive and negative.

A school may achieve higher results than another simply because a larger proportion of its families use private tutoring. A school's results may improve significantly because a higher proportion of families engage private tutoring. This says nothing about the quality of teaching and curriculum at the school.

School results are a selective measure of education

Comparing and judging schools on their test results for literacy and numeracy does not provide a full picture of the work of schools. It fails to consider that education is more than literacy and numeracy, even though these are critical skills. The purposes of school are the intellectual, social, civic, and ethical development of individuals. Literacy and numeracy tests are not necessarily an indicator of the success of schools in pursuing these broader purposes of school.

Different schools often do better in some year levels, in some curriculum areas, and for some schooling objectives. Some schools can do well in some other important areas of learning such as supporting the personal and social development of students, arts and music and science, but not so well in literacy and numeracy. The focus on the results of literacy and numeracy tests fail to take account of the diversity of educational programs offered by schools.

Failure to take account of the contribution schools make in other important areas of childrens' learning may give a distorted and inaccurate view of school quality.

School results are subject to manipulation and rorting

School results may be artificially boosted by being manipulated in various ways. Overseas overseas experience shows that many schools resort to poaching high achieving students from other schools, denying entry to, or expelling, low achieving students, suspending low achieving students on test days, holding back students in grades not tested, increasing use of special dispensations for tests, encouraging students to take courses whose results are not used to compare schools and outright cheating. Extensive academic studies also show that test results are manipulated by schools in various ways to improve their ranking.

Several of these ways of manipulating school results are already being used in Australia. During the recent NAPLAN tests there were many instances of schools encouraging low achieving students to stay home during the tests, leaking of tests beforehand to alert teachers about questions and teachers helping students with answers and changing answers.

The overseas experience with publishing school results and league tables suggests that manipulation of school results is likely to increase in the future as a result of the pressure placed on teachers and principals to improve school results.

School results are subject to statistical error

Considerable uncertainty surrounds the accuracy and reliability of school results because of measurement and sampling error. These errors are inevitable in testing and reporting regimes.

Many technical studies of school results and school league tables demonstrate that chance differences account for a significant proportion of the differences in school test scores. In the

case of gains from one year level to the next or annual changes in the results of a given year level, the margin of error can be exceptionally large. Several studies, including one Australian study, show that the results of up to 80% or more of schools are indistinguishable from the average school outcome. Real differences in school results can be only identified for a small minority of schools.

This level of error wreaks havoc when comparing school results. It is not possible to make reliable comparisons or rankings of schools because they may reflect chance differences in school performance rather than real differences. Such comparisons are mostly identifying lucky and unlucky schools, not good and bad schools.

The Australian Curriculum, Assessment and Reporting Authority has failed to implement the decision of the national ministerial education council that statistical error margins should be published with school results to ensure accurate interpretation. The My School website only reports the margins of error for various cohort sizes taking the test. This is meaningless because the size of the cohort taking each test at each year level in each school is not reported.

Little prospect that the reliability of My School can be improved

There is little prospect that making changes to My School will improve its accuracy and reliability in any substantive way and reduce the scope for parents and the public to be misled in comparing school results.

It is unlikely that sufficiently detailed information can be obtained to accurately measure the socio-economic status of schools. It will also be difficult to obtain information on the detailed ethnic make-up of schools.

Little can be done to counter manipulation and rorting of school results. The “high stakes” attached to published school results and league table rankings mean that rorting and cheating is inevitable. It continues to be a feature of systems that have been publishing school results and league tables for the past 20 years.

At the very least, My School should report the margin of error for each test at each year level for each school so as to more accurately inform parents and the public about school results published on the website. It would ensure compliance with the principles and protocols for reporting school results promulgated by the national education ministers council.

1. Introduction

The former education minister and now Prime Minister, Julia Gillard, claims that publishing school results is needed to help parents choosing a school for their children. For example:

> You'll then be able to compare that school to other schools in your local community which is important if you move there with a few kids you want to see how all the schools go to make a choice about where your child should go. [Gillard 2009b]

The former Prime Minister has said:

> ...what parents want is in a particular geographical area, just say within a wider suburban area, they would like to be able to know how one school is going against the other and we think that's a fair thing too. [Rudd 2008b]

However, comparing school results is unlikely to provide parents and the public with an accurate measure of school quality. Raw school results can mislead parents and others about the comparative quality of schools because:

- Differences in school composition affect school results;
- Many other factors outside schools influence school results;
- They are a selective measure of education;
- They are subject to manipulation and rorting; and,
- There may be significant statistical errors on school test results.

These factors may lead parents to choose a school of lesser quality than its results indicate. Some schools may be wrongly recognised as outstanding while others are identified as unsuccessful for reasons that are outside the control of schools.

These factors also make it difficult to identify effective school practices. Decision-makers and schools may be misled in recommending and adopting particular educational programs. Education practices and programs could be falsely identified as successes while successful programs in reality are ignored or even falsely condemned. Action taken to assist less successful schools may appear more effective than it is in practice.

2. Differences in school composition influence school results

Differences in the student composition of schools by socio-economic, ethnic, Indigenous, gender and disability all influence school results and cause misleading comparisons between schools.

2.1 Differences in socio-economic composition

School results are significantly determined by the socio-economic background of school communities. Many studies show that low family income, low parental educational achievement and low skills are strongly associated with poor education outcomes while high family income, high parental educational achievement, professional and managerial occupations are strongly correlated with high education outcomes [for example, see Haveman & Wolfe 1995; Sirin 2005; Berliner 2009; Skolverket 2010; Xia 2010].

This is reflected in school results in Australia where there are large differences in student achievement by socio-economic background. According to the latest results of the OECD Programme for International Student Assessment (PISA), nearly 25% of 15 year-old students from low income families in Australia do not achieve expected international proficiency standards [Thomson & De Bortoli 2008a]. In 2006, 22-23% of low socio-economic status

(SES) students did not achieve international proficiency standards in reading, mathematics and science compared to only 5% of high SES students. Thus, the proportion of low SES students not achieving expected levels is about 5 times that of high SES students.

In contrast, the proportion of high SES students achieving the highest proficiency levels is about 5 times that of low SES students. In 2006, only 4% of low SES students achieved the highest reading proficiency standard compared with 21% of high SES students. In mathematics, the respective proportions were 6 and 29% and in science it was 6% compared to 26%.

On average, 15 year-old students from low SES families are two years or more behind high SES students. In 2006, the differences in average score points between low and high SES students in reading, mathematics and science were 84, 78 and 87 respectively

The proportion of low SES students achieving below the OECD average is about 2½ times that for high SES students. In 2006, 53-55% of low SES students achieved below the OECD average in reading, mathematics and science compared to 22-24% of high SES students. [Thomson & De Bortoli 2008b].

National data on retention rates shows that the drop-out rate before Year 12 for low SES students is double that of high SES students. In 2008, 42% of students from low SES families failed to complete Year 12 compared to 23% of students from high SES families [MCEETYA n.d: Table 34].

A recent report of the NSW Auditor-General shows that one or two in every 10 low income students are below minimum State standards in literacy and numeracy compared to one or two in every 100 high income students [Audit Office of NSW 2008]. In 2007, 11% of Year 3 students in South Western Sydney and 9% in Western Sydney, both low income regions, were below minimum literacy and numeracy standards in 2007 compared to 1-2% of students in Northern Sydney which is a high income area. The achievement gap was huge for disadvantaged schools where 20% of students were below the Year 3 minimum standard in literacy and 15% were below the numeracy standard.

A report by the Victorian Auditor-General found that the achievement gap between students from low- and high-SES schools was wide at all year levels for both literacy and numeracy [Auditor-General, Victoria 2009]. Students from low-SES schools were up to a year or more below the achievement level of their counterparts from high-SES schools for both literacy and numeracy. This achievement gap widened as students progressed through school from Years 3 to 9. In Year 9, the gap represented 15 months of learning for both literacy and numeracy.

Thus, differences in school results may simply reflect differences in student composition. High test scores or high league table rankings may reflect more the privileged family background and resources of that school community rather than the quality of teaching and the education program. Parents who choose schools on this basis may therefore be misled about school quality.

2.2 Differences in ethnic composition

When students from all ethnic backgrounds, or from non-English speaking backgrounds, are considered as a group they achieve at similar levels to students of Australian-born parents

[Cresswell 2004; Cobbold 2009]. However, there are significant differences in student achievement between different groups of students and differences in school results may reflect differences in ethnic composition rather than differences in school quality.

There is evidence that the average results of Lebanese students and Pacific Islander students are well below those of Chinese students. A study of Year 10 Certificate results for English, Mathematics and Science in three high schools in south-western Sydney found substantial differences between the results of students from different ethnic backgrounds [Suliman & McInerny 2006]. It found a much higher percentage of Lebanese students were achieving in the lower grades in all subject areas than Chinese and Vietnamese students. A much lower percentage of Lebanese students achieved the top grades than Chinese and Vietnamese students. The average results of Lebanese students were significantly below those of Chinese and Vietnamese students.

An analysis of data from the Longitudinal Surveys of Australian Youth shows that achievement in literacy and numeracy in Year 9 varies widely among students from language backgrounds other than English [Marks & McMillan 2000]. When ethnic background was measured by nine categories relating to father's country of birth, some ethnic groups showed higher Year 9 achievement levels than those students with fathers born in Australia, while students from other groups showed lower Year 9 achievement levels. This finding was confirmed by further analysis of university entrance scores. It found that students classified as Asian performed substantially better than students whose fathers were born in Australia when adjusted for socio-economic background [Marks et.al. 2001]. On the other hand, the mean scores of students with Middle Eastern, North African and Pacific Islander ancestries were significantly below those of Asian students.

A study carried out by the Australian Council for Educational Research (ACER) for the National Education Performance Monitoring Taskforce of the Ministerial Council for Education, Employment and Youth Affairs (MCEETYA) shows that average numeracy test scores for Year 9 students from the Middle East, North Africa and the Pacific Islands are significantly below those of Asian students [Ainley et.al. 2000].

A later ACER report which analysed the 2000 PISA results for Australian immigrant children found that average reading and mathematics scores for Chinese and other Asian language students were significantly higher than for students with Middle Eastern home languages [Cresswell 2004]. In terms of proficiency levels, nearly 50% of Chinese students and 29% of Other Asian students achieved at the top two reading levels compared to 15% of Middle Eastern students. In contrast, only 9% of Chinese students did not achieve expected minimum standards compared to 24% of Middle Eastern students.

Thus, differences in school results may reflect differences in ethnic composition rather than differences in school practices. For example, schools with a high proportion of students of Asian origin are likely to achieve much better results than those with a high proportion of Middle Eastern or Pacific Islander students. Parents and others comparing the results of these schools could therefore be misled about the comparative quality of teaching and curriculum in these schools.

2.3 Other differences in student composition

Differences in school results are also influenced by the proportion of Indigenous and students with disabilities.

Schools with a high proportion of Indigenous students are likely to have lower average results than schools with a high proportion of non-Indigenous students. About 40% of 15 year-old Indigenous students do not achieve expected international proficiency standards compared to 13% of all Australian students [Thomson & De Bortoli 2008a].

Similarly, many schools have higher proportions of students with disabilities than others. Schools with higher proportions of students with disabilities participating in tests may have lower results than other schools.

2.4 Policy contradictions

There is a deep contradiction in the Government's approach to reporting school results. It has acknowledged that comparisons of school results can be misleading but it has gone ahead with publishing school results.

The new Prime Minister has stated that she is opposed to "simplistic" league tables because they lead to unfair comparisons between schools facing with vastly different circumstances. For example:

> But what is too simplistic is just comparing a school in the richest suburb in the country with an outback school serving predominantly Indigenous kids and comparing the two of those. That's not going to tell you much. We all know that the school in the richest suburb is going to do better. [Gillard 2008]

> I understand that league tables based on raw test scores can create a misleading picture and make the jobs of principals and teachers that much harder. [Gillard 2009a]

Yet, these comments apply just as much to comparisons of school results published on My School as they do to league tables.

The former Prime Minister was equally adamant about how unfair and misleading it is to compare the results of schools facing different circumstances:

> Simplistic league tables don't really tell us how well a school is performing. They don't tell us about the student population that the school started with – and its level of educational advantage. Everyone understands why a private school on Sydney's north shore might do better than a comprehensive government high school in the outer suburbs. [Rudd 2008a]

He also stated that he didn't want to see the likes of Geelong Grammar compared with a school such as Nambour High, where he went to school because "it's got a different set of, shall we say 'challenges', than Geelong Grammar" [Rudd 2008b].

Yet, this is precisely what My School does – it compares schools comprised of high SES intakes with schools having low SES intakes. For example, its local area page for Geelong Grammar – one of the wealthiest schools in Australia – compares its results with government and Catholic schools in Corio and Norlane, which are amongst the most disadvantaged suburbs in Australia and are even more disadvantaged than Nambour. Geelong Grammar has 1% of its students in the bottom quarter of the My School socio-economic index compared to over 90% for many other schools in the area including Corio PS, Corio South PS, Corio West PS, and St. Thomas Aquinas Catholic School. My School includes many other unfair and misleading comparisons between some of the most advantaged and disadvantaged schools around Australia.

The fact is that the My School website permits comparisons of schools without regard to their composition. The Government claims that context for comparing school results is provided by the so-called "like school" comparisons on My School. However, none of the above problems are rectified by "like school comparisons" because they fail to consistently compare like with like [Cobbold 2010].

3. Other outside factors influence school results

Many other factors outside the control of schools also influence a school's results. These include student absenteeism, the extent of parent involvement in learning at home, the extent to which students change schools, private tutoring, school size and funding. and the extent to which students are engaged in private tutoring.

For example, some schools may have lower results because they have a high proportion of students who often change school rather than because their teaching quality and curriculum are inferior. Studies show that student mobility in Australia can be high. A Queensland study has shown that 16% of all primary school students moved school 2 or 3 times in five years and 4% moved school 4 or more times [Simons et.al. 2007]. Another Queensland study also shows very high student mobility in some schools [Hill et.al. 2009]. Other data shows that nearly 40 per cent of students in the Northern Territory change schools in any one year, apart from those who go on to a higher stage of schooling [Dunn 2009]. Even in a high income city such as Canberra, there are schools where annual turnover is over 30% at times.

While student mobility is higher amongst low SES families than for high SES families, it is also significant for high SES families. For example, the Queensland study found that nearly 30% of low SES primary school students and 13% of high SES students changed school 2 or more times in five years.

Many studies show that students who move school often tend to have lower average results than students who remain at the same school; although there is also evidence that it has little effect [Sorin & Iloste 2006; Simons et.al. 2007]. A recent syntheses of research studies on the effects of school mobility on achievement and dropout rates found that the large majority of studies conclude that children who move school 3 or more times have significantly lower reading and mathematics achievement and are more often to drop out of school [Reynolds et.al. 2009].

Thus, schools could have large differences in average literacy and numeracy test scores because of differences in the proportion of students who often change schools. It also begs the question of the extent to which NAPLAN results can be credited to a school when a significant proportion of its students are only recent enrolments.

School size is another factor influencing school results. School results and league table rankings may significantly distorted by the results of a few students in small schools. Studies show that small schools are much more likely to report large changes in average results from one year to the next, both positive and negative [Kane & Staiger 2001; Linn & Haug 2002; Wu 2009b].

Another much ignored influence on school results is private tutoring. The use of private tutoring by families in Australia is so extensive that it has been called a 'shadow education system' [Watson 2008]. Schools may have different proportions of students engaged in private tutoring. In addition, changes in school results from one year to the next may be

influenced by changes in the proportion of families who use private tutoring. If a higher proportion of families engage private tutoring in any one year a school will receive a boost to its measured performance compared to another school even though there was no change in teaching effectiveness during the year.

4. Published school results are a selective measure of education

Comparing and judging schools on their test results for literacy and numeracy does not provide a full picture of the work of schools. It fails to consider that education is more than literacy and numeracy, even though these are critical skills. The purposes of school are the intellectual, social, civic, and personal development of students. Literacy and numeracy tests are not necessarily an indicator of the success of schools in pursuing these broader purposes of school.

Different schools often do better in some year levels, in some curriculum areas, and for some schooling objectives. Some schools can do well in other important areas of learning such as supporting the personal and social development of students, arts and music and science, but not so well in literacy and numeracy. Focus on the result of literacy and numeracy tests fail to take account of the diversity of rich educational programs offered by schools. Failure to take account of the contribution schools make in other important areas of childrens' learning may give a distorted and inaccurate view of school quality.

This is not to suggest that NAPLAN assessments should be expanded to give a more comprehensive coverage of student learning. These measures would also be affected by factors outside the control and influence of schools. In addition, prohibitive costs would be incurred in extending standardised assessments to all learning areas and other year levels. Such resources would be better devoted to directly improving student learning.

5. School results are subject to manipulation and rorting

An inevitable outcome of reporting school results is that they are corrupted by the strategic responses by schools that are now commonplace overseas where standardised tests are used to rank, reward and punish schools. Corruption of test scores occurs in two broad ways. One is to undermine the sample on which the test is based by finding ways to exclude lower achieving students. The other is to distort the measure itself by cheating, increasing the use of special dispensations for tests or by increasing the time devoted to areas that are subject to testing.

Competition for higher rankings forces schools to "play the system" to show improvement even where there is none. Playing the system is the quick route to better results.

An example of the absurd lengths that school systems will go to fudge their results comes from Texas. New rules introduced last year allow schools to count students who failed the Texas Assessment of Knowledge and Skills (TAKS) as passing, as long as a complex formula shows that those students are predicted to pass in a future year [*Dallas Morning News*, 5 July 2009].

Under the new arrangements, if a student fails the Year 7 math TAKS, the school can use a statistical formula developed by the Texas Education Agency to predict whether that student will pass the math test in Year 8. The formula considers the student's math and reading TAKS scores, plus the average math TAKS score at the school. If the student is predicted to pass, the school gets to count her/him as actually passing in Year 7 even though she/he really failed.

The new system is foolproof against failure. A school never has to go back and compare the predicted performance with the actual performance in Year 8. A school can record a Year 7 student who failed the TAKS as a pass if the student is projected to pass in Year 8, but it is not penalized if that student does not pass in Year 8 as predicted. Instead, the model looks ahead again to predict whether the student will pass the Year 11 TAKS.

The result has been that hundreds of schools received a higher rating on the TAKS. The number of "exemplary" schools (the highest rating) more than doubled from 1,000 in 2008 to 2,158 in 2009 [*Houston Chronicle*, 20 June 2010]. Without the statistical projections that some failing students would later pass, the increase would have been only 44 schools. At the other end of the spectrum, the number of "unacceptable" schools increased by 43, from 202 to 245. But without the use of the statistical projections, the real increase was 401, or almost 10 times the adjusted number.

Apart from such spectacular examples of manipulation of school results, overseas experience shows that many schools resort to poaching high achieving students from other schools, denying entry to, or expelling, low achieving students, suspending low achieving students on test days, holding back students in grades not tested, increasing use of special dispensations for tests, encouraging students to take courses whose results are not used to compare schools and outright cheating. Extensive academic studies also show that test results are manipulated by schools in various ways to improve their ranking.

5.1 Reduced participation of low achieving students in tests

Schools resort to several ways to exclude low achieving students from tests used to report and rank school performance. They include:

- Increasing classification of low achieving students as special education students who can be exempted from tests;
- Increasing suspensions of low achieving students or encouraged absences on test days; and
- Holding students back in non-tested years.

Other ways of reducing the participation of low achieving students in high stakes tests include encouraging students to drop-out or leave a school before key graduation exams/assessments and encouraging them to take subjects that are not included in the ratings tests. These practices are not discussed here because they are not relevant to the year levels for which NAPLAN is conducted.

Research evidence

A number of academic studies demonstrate that schools respond to accountability pressure by differentially reclassifying low-achieving students as special education so that their scores will not count against the school in accountability systems. Several studies of the Texas accountability system have found that schools were exempting students strategically to improve their rating. In a particularly comprehensive treatment of individual schools' incentives to exempt students, one study found that proximity to a higher accountability rating increased the percentage of students exempted by 11% in consecutive years [Cullen & Reback 2006]. The exemptions were driven by more aggressive special education placements and absenteeism. The study also found that when the performance of Hispanic and African-American students would keep schools from achieving a higher rating, exemption rates increased for these groups by 7 and 14% respectively.

A longitudinal study of one Texas school district found a substantial increase in student exclusions from the state tests in elementary schools through special education and language exemptions [Heilig & Darling-Hammond 2008]. Higher proportions of black and Hispanic students were exempted and they were lower achieving students.

This study also revealed the interesting practice of "grade skipping", in which students stayed in 9th grade for 2 years or more and then suddenly reappeared in the 12th grade. This practice had two benefits for schools. First, by skipping 10th grade, students did not take the state test in the year that it counted for school accountability ratings. Second, by showing up in 12th grade, they contributed to a more favourable rating where school progression is examined as the proportion of 9th graders who appear in 12th grade 4 years later.

A study of the Houston school district found that lower achieving students are much more likely to be exempted from the state tests than high achieving students [Jennings & Beveridge 2009]. The exemptions caused a significant increase in school pass rates. About 26% of all schools in Houston would have achieved a lower accountability rating in reading if the students had not been exempted and about 22% would have achieved a lower rating in mathematics. About 60% of schools rated "exemplary" would have achieved a lower rating.

A detailed case study of one Texas school demonstrated how the school attempted to influence its test rating by removing potentially low-scoring students from the accountability subset by having these students tested to determine whether they qualified for special education [Booher-Jennings 2005. Teachers referred more students for special education testing in response to the state accountability system.

The introduction of high stakes testing in Florida led to about a 50% increase in the rate that students from low-income families were exempted from test-taking due to special education classifications [Figlio & Getzler 2006]. Exemption rates were also much higher in high poverty schools than more affluent schools, indicating a greater tendency for low performing schools to exclude students from tests as a way of increasing their reported outcomes.

A study of Chicago public schools found that schools increased the proportion of students classified as special education so that they were not required to sit the national tests [Jacob 2005]. The largest increases in special education placements occurred for low achieving students in low achieving schools. A study using Ilinois state test data found that schools strategically chose some students not to take the test in order to increase their pass rates [Lemke et.al 2006]. It found that schools with low pass rates allowed more of their students not to take the test and strong statistical evidence that having fewer students take the test was associated with higher pass rates.

Schools may also use selective disciplinary policies to change their testing pool to improve their pass rates and school ratings.

A study of the introduction of a high-stakes testing regime in Florida has showed that that schools differentially suspend students at different points in the testing cycle so as to alter the composition of the testing pool [Figlio 2006]. While schools tended to assign harsher suspensions to low performing students than to high performing students, the gap grew substantially during testing periods and only in the grades tested. Schools reduced their suspension penalties for higher achievers students in the grades subject to high stakes tests and increased suspension penalties for low achieving students in the same grades at this time.

High achievers received shorter suspensions and low achievers higher suspensions at test time and the week immediately preceding the test compared to other periods during the year. The study concluded that while the changes in suspension during test time had a small impact schools' ratings this small impact may be significant for schools on the margin of rating categories.

There is also evidence of schools in other US states using disciplinary measures to exclude low achieving students from school on test days [Nicholls & Berliner 2007]. Some schools exclude low achieving students by enforcing zero-tolerance discipline policies and expelling students for attendance problems [Heilig & Darling-Hammond 2008].

There is also some evidence of such practices from England, where there was a threefold increase in the number of permanent exclusions from schools after the introduction of league tables [West & Pennell 2000]. One study of exclusion practices in England found that while the factors contributing to student exclusion from school are multiple and complex, increased test-based school accountability increased the pressures and incentives to exclude low achieving students [Rustique-Forrester 2005].

Another way for schools to influence the test taking pool by excluding low achieving students is to hold them back from the tested grades. By doing so, schools give these students an additional year of learning before moving to the next grade and facing the high stakes test. For example, a study of Chicago public schools found evidence of increased use of grade retention to give students an additional year of learning [Jacob 2005]. Another study found that schools in one Texas school district were increasingly using grade retention as a way of influencing test results [Heilig & Darling-Hammond 2008]. Such practices were also found in a case study of a Texas school [Radigan 2007].

Evidence in Australia

The manual of procedures issued for NAPLAN is explicit about exemption of students from the tests. All students are encouraged to participate in the tests but students can be exempted on three grounds: students with a significant intellectual disability; students newly arrived from overseas and at the request of parents. In each case, the decision lies with parents and they have to apply for exemption.

These provisions have created loopholes for schools to exploit and they were used all around Australia during the NAPLAN test period this year. In some instances, they were pushed open even more. Schools used these provisions to take pro-active action to encourage parents of low achieving students to apply for exemptions or to exercise their option not to have their children participate in the tests.

There were many media reports of parents of low achieving students in Queensland and Victoria being told by schools to keep their children home from school over the three days of NAPLAN testing.

The Melbourne *Herald-Sun* said that dozens of parents and teachers had contacted the newspaper telling of schools that had put pressure on children to stay home and not "drag down" the school averages [*Herald-Sun*, 12 May, 13 May]. One Melbourne teacher told *The Age* he was aware of parents of failing students being told during parent interview nights there was no educational benefit for their child to sit the NAPLAN test [*The Age*, 13 May]. *The Herald Sun* was told four grade 3 students at a school in the Loddon Mallee region were

told not to sit the NAPLAN test because it might bring down the school's results. Their parents signed forms exempting them from the test and they spent the day with grade 4 and 6 students, who did not do tests.

The President of the Queensland Teachers Union told *ABC News* [11 May] that he was aware of several cases of schools encouraging students to stay home on the test days.

> It's leading to all sorts of unnecessary practices in schools. I have heard of it happening in schools where they've deliberately taken a stance that they don't want below-average students doing the tests and dropping their scores, which probably shows less of an understanding of what the NAPLAN tests are designed to do rather than anything else.

A parent at Mount Cotton state primary school in Queensland said she was told her son was exempt because of an intellectual disability and he would either be put in a grade 2 class while the tests were on or she could keep him at home [*ABC News* 11 May, 13 May; *The Australian*, 12 May].

The father of a struggling year 7 student at Vermont Secondary College in Victoria said his son was told he did not have to sit the NAPLAN tests and that his wife was contacted by a school co-ordinator on the night before the tests started and told their son did not have to sit the test.

> My son is a C and D-grade student - he doesn't receive A's. I couldn't believe it … schools are obviously trying to get themselves to look better than they actually are and that's wrong. You can't fudge the figures - it's fraudulent. [*The Age*, 12 May]

The parents of a Year 9 student at Leongatha Secondary College were contacted by the school half an hour before the first test was due suggesting that their son not sit the tests [*Herald-Sun*, 13 May].

These incidents demonstrate that many schools have not followed the spirit or the letter of the NAPLAN administrative guidelines. Schools have taken a pro-active stance to encourage some students not to sit the tests rather than to encourage all students to participate and to leave it to parents to decide whether to seek exemption for their child.

One school principal defended the school's action in contacting parents to withdraw their children on the grounds that the student concerned "may find doing the test sitting for that length of time frustrating" [*ABC News*, 13 May]. The President of the Victorian Association of State Secondary Principals said:

> There's not much point in a child with an intellectual disability sitting there doing a test where they could do five minutes of the test and they're sitting there for an hour ... nobody's gaining out of that. [*ABC News*, 13 May]

The acting principal at Brauer College in Warrnambool defended asking the parents of some students not to sit the tests because the students may find the tests daunting:

> There were students who we thought would be distressed by doing the tests and they are for the most part students in the disability program or with reading ages that are hugely below expected....We then contacted their parents and if the parents wanted them excluded then that's what we did." [*ABC News*, 13 May]

These are not grounds for exempting students. They are not provided for in the administrative guidelines for NAPLAN and amount to bending the rules. If such action is not stamped out

by education officials more and more schools will resort to the loophole to improve their school results. More and more schools will pressure parents of low achieving students to withdraw their children from the tests as a way of improving school results.

The response of the then Federal Education Minister, Julia Gillard, to these incidents verged on the complacent. She told *The Age* that school participation rates in the national literacy and numeracy tests would be published on the My School website: "If there is an unusually low number of children participating in a school, that will prompt questions and it will be investigated," she said [*The Age*, 13 May].

Such investigations are likely to prove ineffectual. There will be variations in participation rates in NAPLAN from year-to-year, and it will be impossible to determine whether any reductions are due to schools initiating exemptions or whether they are due to parents deciding on their own to seek exemptions. Changes in participations rates will also be affected by differences in the incidence of illness amongst children from year-to-year. Also, high rates of mobility between schools may also lead to changes in participation rates if a new influx of students has a higher rate of absenteeism.

Schools will be able to strategically exclude low achieving students without triggering a large reduction in participation rates which would attract an investigation by from education officials. A small increase in a school's exemption rate may produce a large impact on school results if it is targeted at the lowest achieving students.

5.2 Controlling school admissions to enrol higher achieving students

An easy way for schools to improve their performance is to replace low achieving students with those who generate better test results. This involves actively poaching, or 'cream skimming', high achieving students from other schools while denying access to low achieving students.

There is abundant evidence from overseas of schools selecting their students to maintain or improve their league table position. For example, a favourite strategy is to use formal and informal enrolment criteria to "cream skim" or "cherry pick" students most likely to achieve good results. Formal processes are used to select students for 'ability', 'aptitude' or 'motivation' and informal or 'covert' selection processes are used to discourage 'undesirable' students. The latter include the less 'able', children with emotional or behaviour problems, students from low socio-economic communities, children with learning difficulties and other special needs.

In England, the misuse of school admission procedures to maximize school results has been a major ongoing issue since the introduction of league tables. Various methods have been used to select certain groups of students and exclude others, such as giving priority to the children of employees, former students, those with a family connection to the school, and selecting a proportion of children on the basis of aptitude/ability in a subject area(s) or on the basis of general ability [West et.al. 2004]. Many schools also required parents to provide supplementary information unrelated to the school's admissions criteria, such as their occupation, whether the family lived in a hostel or bed and breakfast accommodation and whether parents had refugee status. Many of these practices continue despite government efforts to stamp them out [West et.al. 2009].

A recently published study of public sector religious secondary schools in London has found that selective 'élite' schools appear to 'select out' low income religious families, thereby displacing them to religious schools with a less affluent composition [Allen & West 2009]. It identified a range of different admissions criteria and practices used by the socially selective schools including school-administered banding, aptitude tests, tests of religious or denominational commitment, primary school references and others that may have contributed to the under-representation of lower ability pupils. Moreover, the hierarchy of schools that has developed also tends to dissuade many parents from even applying to the more selective schools because they believe they will not meet the selection criteria.

> ...schools' admissions criteria and practices are important determinants of which pupils are offered places, whether on account of schools attracting applications from certain parents or in terms of the admissions process itself. [Allen & West 2009: 19]

A recent study by academics from the London School of Economics found that many schools are still covertly selecting students on the basis of academic ability despite it being outlawed by government regulations [Noden & West 2009]. It found that several schools are "gazumping" each other to attract the best students and principals are employing underhand tactics which are outside the official guidelines for school admissions. It said that a stricter admissions code introduced by the UK Government to make admissions fairer was not enough to stop schools poaching students from one another. It said that it was "not difficult to find schools that fall foul of the code".

The dubious practices include contacting parents to persuade them to reject offers from more highly preferred schools and ranking students on waiting lists according to their own criteria rather than the official rules which give priority to disadvantaged children and those with special needs. One school selected students on the basis of how near their homes were to a building half a mile from the school in an attempt to upgrade its student intake. Another school removed the sibling rule from its admissions criteria so as to break the link with the social composition of the existing student cohort.

While the new admissions code states that it is necessary to improve the chances of more disadvantaged children getting into good schools, it appears that this is being ignored in many instances [*The Guardian*, 8 December 2009]. As a result, students from low income families sometimes can even be left without a place at their local school.

There is also evidence of similar selection procedures in the US. For example, a recent study of a large urban school district in Texas found that a number of schools were refusing to enrol low achieving students in order to protect their school accountability ratings [Heilig & Darling-Hammond 2008].

With the launch of My School, this is likely to become a major issue in Australia with its large private sector. Private schools have the ability to select their students and they already engage in poaching students from other private and government schools. Publishing school results and league tables are likely to lead to even greater selection to protect school reputations and rankings. Such selection processes may also extend to the government sector where some schools are in such high demand that they can effectively discriminate amongst applicants for enrolment.

Already, there is some evidence of this. According to the *Herald-Sun*, the spokesman for Independent Schools Victoria said it was common practice to use NAPLAN tests for

assessments for entry to private schools [12 May 2010]. The report cited the example of Mentone Grammar School whose website states:: "Parents are requested to bring along to the interview copies of your children's last two school reports, NAPLAN test results and any other relevant information or assessments you might have."

One mother told *The Courier-Mail* during the NAPLAN tests that she had practised the NAPLAN tests with her daughter to help her do well because high schools were now deciding enrolments using NAPLAN results [12 May 2010]. According to the newspaper, many parents told Brisbane talkback radio their children's NAPLAN results had been required when they sought to enrol their children in a private school.

It has also been reported that private schools around Australia are considering introducing pre-entry tests for students which could be used “to weed out and exclude potentially poor performing students” [*Adelaide Advertiser*, 26 March 2010].

5.3 Cheating

School results can be manipulated by cheating. The high stakes of reputation and status associated with reporting school test results create incentives for teachers to help students with answers in tests and change answers. An academic study has shown that teachers are more likely to cheat when faced with more accountability pressure [Jacob & Levitt 2003].

Cheating can take various forms. Pre-test cheating occurs where teachers alert students to test questions or they use access to the test to prepare students for particular questions. Cheating during tests occurs where teachers directly help students with their answers during tests or allow students to cheat. Post-test cheating involves teachers changing answers before submitting test sheets for marking.

Cheating incidents overseas

Cheating is an endemic problem in the US and England where school results are published and has been so since publication of school results and league tables were introduced [for a survey of cheating incidents a decade ago see ACT P&C Council 2000]. In the twenty years or more since school tests results have been published in England and the United States, no government has been able to stamp out cheating. Indeed, it appears to have increased rather than diminished.

Cheating incidents have been reported in many US states in recent years. For example, last year a survey of public school teachers in Chicago by the *Chicago Sun-Times* [29 August 2009] and the teachers’ union revealed that one-third of all teachers had been pressured in the last year by principals and boards to change student grades. Twenty per cent said that they had actually raised grades under this pressure.

Earlier this year, one of the largest cheating scandals ever in the United States was reported involving over 400 schools in the state of Georgia [*Atlanta-Journal Constitution*, 11 February 2010; *New York Times*, 12 February 2010]. Some 191 schools are under investigation for tampering with test answers and another 178 schools are subject to increased monitoring of their tests. The main focus of the investigation is Atlanta, where 70% of all elementary and middle schools face investigation. At one school, for example, averages of 27 of 70 answers on each fourth-grader’s math test were changed from wrong to right in one classroom. At another an average of 26 of 70 answers on the fifth-grade math test were erased and corrected. More than half the classes at 27 schools were flagged, and at four Atlanta schools more than 80 percent of the classes were flagged.

Experts said it could become one of the largest cheating scandals in the era of widespread standardized testing in the US. Gregory Cizek, Professor of Educational Measurement and Evaluation at the University of North Carolina, told the *Atlanta Journal Constitution* (11 February) that the extent of the suspicious answer changes is stunning. He has studied cheating for more than a decade, but said he didn't know of another state that has detected so many potential problems.

The *New York Times* [10 June 2010] recently reported that cheating by teachers and principals was being investigated in Indiana, Massachusetts, Nevada, Texas, Virginia and other states as well as in Georgia.

The Texas school accountability system was the forerunner of many school reporting systems in the US, including the No Child Left Behind Act. Ever since it was introduced, it has been plagued by cheating of school results which has continued to this day. Just recently, the *Houston Chronicle* (25 May) reported that a Houston school district found evidence that teachers had changed some fifth-grade students' answers on the Texas Assessment of Knowledge and Skills exam and helped students correct wrong answers. Other evidence shows that some teachers had access to the exam answer sheets and possibly the exam questions before test day. A principal, deputy principal and three teachers have resigned as a result of the investigations.

Experts say that cheating is increasing as the stakes over standardized testing ratchet higher, including, most recently, taking student progress on tests into consideration in teachers' performance reviews. A data forensics expert told the *New York Times* [10 June 2010] that cheating was on the rise in the US. "Every time you increase the stakes associated with any testing program, you get more cheating," he said.

Others say that what is revealed is just the tip of the iceberg. For example, one expert, emeritus Professor Tom Haladyna from Arizona State University, told the *Atlanta Journal-Constitution* following other incidents of cheating in Georgia last year that: "It's just the tip of the iceberg, I think. The other 80 percent is being hidden" [21 June 2009].

Education departments across the US are now spending millions of dollars in trying to monitor and deter cheating designed to bolster school results. Several states have hired test security companies to do audits of tests to check for cheating.

Many incidents of cheating have also been reported in England since league tables were introduced nearly 20 years ago [ACT P&C Council 2000]. They continue to turn up. Between 2002 and 2005, nearly 250 teachers were investigated for offering their students advice on how to answer questions in national tests [*Times Educational Supplement*, 1 September 2006].

A couple of years ago, a *BBC News* investigation exposed widespread cheating by teachers in exams in England in order to raise school ratings on league tables [27 July 2007]. It cited the results of small survey in which two-thirds of teachers said they personally help students "more than is appropriate" in order to improve exam results. Teachers told the BBC that cheating was a taboo subject which everyone knows about it but nobody wants to speak out. The former head of the Office for Education Standards told the BBC that cheating by teachers is so extensive that the league tables used by parents to differentiate between schools have become unreliable.

Last year, it was reported that students at 70 schools had theirs test results annulled or changed because of cheating by teachers [*Daily Telegraph*, 28 July 2009]. In some cases, students were deliberately coached during the exam in order to reach the correct answer. Earlier this year, a teacher was banned for two years after helping students with an exam [*Times Educational Supplement*, 1 January 2010].

Cheating in Australia

Already allegations of cheating have been made in Australia in only the first NAPLAN tests since My School became operational. Several allegations were made during the NAPLAN tests in May this year in several states. They involved the leaking of tests beforehand to alert teachers about questions and teachers helping students with answers and changing answers. An innovative method was leaving posters on the walls of classrooms that students could refer to while doing the tests.

The Western Australian State School Teachers Union said it had evidence that schools around Australia opened the NAPLAN tests early and prepared their classes accordingly [*ABC News*, 13 May]. It called for an inquiry to claims that some schools cheated on national literacy and numeracy tests to boost their ranking [*ABC News*, 12 May].

The head of a prestigious Perth boys' school warned that the possibility of cheating on national literacy and numeracy tests makes them too easily corruptible to be an effective tool to measure school performance. The Christ Church Grammar School headmaster said rumours had circulated that teachers who received test papers several days before students sat the tests were using their knowledge of the contents to prepare their classes. [*West Australian*, 12 May).

> In the week of the tests other rumours exist of teachers being in possession of the test papers prior to the day of their implementation and using this knowledge as last-minute test preparation with classes....It's a totally corruptible tool the Government has now put in place.

The WA Education Department said it was investigating two allegations of cheating [*ABC News*, 15 May]. It subsequently found no evidence of cheating [*ABC News*, 20 May].

NAPLAN test materials were allegedly leaked in Queensland in the lead-up to the tests. Evidence of alleged cheating was cited in a letter sent to education authorities one week before the tests started which warned that some of the writing test items had been leaked for students to practice before the tests [*Courier-Mail*, 12 May]. The *Courier-Mail* also reported that a teacher had contacted the newspaper alleging items from the writing test had been leaked to schools.

The Queensland Department of Education said it was aware of three possible breaches of the testing process [*ABC News*, 13 May]. It was also reported that a question from the spelling, punctuation and grammar test was leaked in NSW [*The Australian*, 12 May].

The New South Wales Department of Education allowed some schools to schedule their tests over two weeks which created the potential for students and teachers in schools which did the tests in the first week to pass on questions to those in schools doing the test in the second week [*The Sydney Morning Herald*, 11 May].

A teacher at a South Australian state school, St. Leonard’s Primary School, was suspended after admitting to altering students' answers on the tests. One report said the teacher was

caught making the changes on the year 7 tests by another member of staff [*ABC News*, 13 May]. Another reported students saying that the teacher stood over them instructing them to erase the answers because "they weren't neat enough" and then indicated which answers to remark [*The Advertiser*, 14 May].

Two other allegations of cheating in Adelaide schools were also made [*ABC News*, 17 May; *The Australian*, 18 May; *Adelaide Now*, 18 May]. A teacher at Elizabeth Vale Primary School is accused of providing what state Education Minister Jay Weatherill described as "inappropriate assistance" to Year 3 and 4 pupils during NAPLAN tests. The teacher has been stood down but denies any misconduct. In the other case, a Year 5 class at the prestigious Catholic school, Rostrevor College, was allegedly given advance notice of the writing test topic. The Catholic Education Office confirmed that a teacher has been stood down, pending an investigation.

A father of a student at Brisbane State High School told the ABC that grade nine students were not supervised during a lunch break in the middle of the tests and were able to compare answers and change them after the break [*ABC News*, 14 May].

> The kids that didn't know the answers to the difficult questions went and saw all the smart kids, got the answers, then they were allowed back into the hall and they were able to do part two....A lot of the kids were updating the first part of the exam with the answers that they got from the smart kids at lunchtime.

It was also alleged that a similar incident occurred at All Hallows School, a private school.

A teacher at the Melbourne school, Debney Park Secondary College, is being investigated for helping a student with a question during the Year 9 numeracy test [*Herald-Sun*, 19 May]. It was reported that the teacher suggested the student change the answer and other students reported it to senior staff at the school.

An innovative method of cheating was also revealed in southern NSW and the Gold Coast. *The Canberra Times* [19 May] reported that students at a NSW school in the Canberra region were told by teachers to look at grammar and spelling charts pinned to the walls of the classroom during last week's NAPLAN tests. A parent at the school claimed that several large laminated grammar charts with explanations and spelling lists were pinned to the walls and written on boards. The parent complained to the school principal and the material was taken down while the tests were in progress. Similar materials for mathematics were also removed before the numeracy tests taken. The NSW Department of Education is investigating the case.

A former teacher at the school told the *Canberra Times* that similar posters had been in place during the national literacy and numeracy tests in previous years:

> I can categorically say that since NAPLAN's inception in 2008, [the school] has given students an unfair advantage over other schools by leaving classroom learning charts on the wall – from spelling to grammar to maths.

A similar case was also reported at Merrimac High School on the Gold Coast [*Courier-Mail*, 19 May]. It was claimed that posters providing basic mathematics information were on display in classrooms during the numeracy tests.

Schools and teachers in Australia now face immense pressure to improve school results and their ranking on school league tables published in the media. School reputations are at stake. Careers are on the line. The Prime Minister, Julia Gillard, has threatened principals and senior staff with the sack if schools fail to lift their results. She wants parents to confront teachers on poor test results.

The Western Australian Primary Principals' Association president said the higher the stakes, the greater the risk that cheating could occur [*West Australian*, 12 May]. Queensland Teachers Union president, Steve Ryan, warned that cheating "is inevitable" because of the way governments are promoting NAPLAN [*Courier-Mail*, 12 May]. Flinders University literacy expert Dr Barbara Nielsen said it was "just a matter of time" before this type of cheating occurred because of the high stakes now involved [*Adelaide Now*, 18 May].

It is not surprising that some succumb to this pressure by cheating. The principal of Rostrevor College, where one teacher has been accused of cheating, said that primary and junior secondary school teachers in his school were feeling "quite significant pressure" because of the NAPLAN tests [*The Australian*, 18 May]. The SA state President of the Australian Education Union said that the teacher at St. Leonard's Primary who admitted cheating felt under pressure because the results were used to assess the performance of schools [*The Australian*, 15 May].

5.5 Extending the school year

Another strategic response to the publication of school results and league tables by some schools is to extend the school year to allow more time to prepare for tests. For example, a recent study of the adoption of state accountability testing in the US state of Wisconsin found that some low-scoring school districts advanced their school start dates to allow their students more time to prepare for exams [Sims 2008].

This too, appears to be a response already being adopted by some schools in Australia. For example, it was recently reported that a prestigious Perth private school has changed its term dates for next year to end Term 1 and start Term 2 a week earlier to better prepare students for the NAPLAN tests [*ABC News*, 9 June 2010].

5.5 Manipulation of school results is a feature of high stakes testing

Everywhere that governments publish school results and permit league tables to be published, schools and schools systems manipulate their results to look better. Australian schools have now joined this strategic response. During the NAPLAN tests this year, there were many instances of schools encouraging low achieving students to stay at home while the tests were on, cheating by changing student answers on tests, leaking questions before the tests, intensive practicing of tests and selective enrolment of high achieving students were all used by some schools to artificially inflate their results.

The incidents were dismissed by the Education Minister as only a small number. However, they are just the beginning. It can be expected that schools will resort to rorting and rigging their results on a grander scale under the pressure to improve or maintain their ranking on school performance tables.

Not only are school reputations under threat as never before, but principals and teachers are being placed under tremendous pressure to improve school results. The president of the Queensland Association of State School Principals said that education officials had told principals to lift their results "at all costs" [*Courier-Mail*, 10 June 2010]. The Australian

Primary Principals Association said that principals in South Australia had been placed under extreme pressure from district managers to improve their test results [*The Advertiser*, 30 March 2010].

All this creates incentives for schools to fudge their results. When testing is simply used as a diagnostic tool, there is no reason for teachers or schools to trick or cheat. This only comes when "high stakes" are attached to the results, such as affecting school reputations and the careers of teachers and principals.

What we saw during the NAPLAN tests was the beginning of a system of fraud – a system in which school results are systematically rorted and rigged. It means that parents and the public will be misled about actual school results. It will not be possible to trust the results posted on My School or the rankings of school league tables as a guide to school quality and progress. Parents and the public will not know which schools are rorting or cheating on their results.

To see the future, one need look no further than the system of school reporting so admired by Julia Gillard – that of New York City and her mentor, Joel Klein. Here fraud and distortion are endemic. Diane Ravitch, Professor of Education at New York University and former US Assistant Secretary of Education, calls it a system of "institutionalized lying" which produces "rigged and fraudulent" results. She says that testing and reporting of school results in the US has become so corrupted that:

> I fear that American education has now entered into a twilight zone, where nothing is what it appears to be, where numbers are meaningless, where public relations and spin take the place of honest reporting, where fraud is called progress. [Ravitch 2009]

6. School results are subject to statistical error

6.1 Measurement and sampling error in test results

Considerable uncertainty surrounds the accuracy and reliability of school results because of measurement and sampling error. These errors are inevitable in testing and reporting regimes [Koretz 2008]. The extent of statistical error in testing varies according to the sample size. As the sample size increases the standard error decreases. Small samples result in larger errors. This is particularly relevant to reporting school results where the years tested include very few students. For example, many schools in Australia have only 25-30 students or less in the years tested under the national assessment and reporting program.

Many technical studies of school results and school league tables demonstrate that chance differences account for a significant proportion of the differences in school test scores. Key US studies show that random factors accounted for almost 15% of the total variation in average fourth-grade test scores for combined reading and maths across North Carolina elementary schools, almost 50% of the total variation between schools in gains in scores during fourth grade, and a massive 73% of the variation in annual changes in fourth-grade scores [Kane & Staiger 2002a; see also Kane & Staiger 2002b]. These results were for schools with an average of 56 students per Year level.

A study of the proportions of students performing at different achievement levels in grades 4-8 in over 400 school districts in Iowa found that sampling error accounted for about two-thirds of the observed variability of estimates of change in proportions from one year to the next [Arce-Ferrer et.al. 2002]. Other sources of error, such as measurement error and

equating error, and intervention effects jointly accounted for about one-third of the observed variability in estimates of change in proportions.

The extent of random errors varies according to school size. A very influential US study found that over various school sizes (for Year levels tested) about 50-80% of changes in average school scores from year-to-year in North Carolina schools was due to random or non-persistent factors [Kane & Staiger 2002a; see also Kane et.al. 2002]. It found that for the smallest one-fifth of North Carolina elementary schools (average of 28 students per Year level), random factors accounted for 20% of the total variation in average fourth-grade test scores, 58% of the total variation in gains in scores during fourth grade, and 79% of the variation in annual changes in fourth-grade scores. For California schools, 86% of the year-to-year changes in test scores for small schools were due to random causes.

One way to gauge the significance of statistical error is to report it as a margin of error, or degree of uncertainty associated with the result. The margin of error is estimated as the range of scores in which there is a reasonably high probability that the true score lies.

Several technical studies have estimated the extent to which school results are different from the overall average (whether for the mean score or the average gain over time) and the size of the uncertainty interval for each school. Intervals that overlap the overall average score indicate no statistical difference in school performance. In the case of gains from one year level to the next or annual changes in the results of a given year level, the margin of error can be exceptionally large.

Several studies, including one Australian study, show that the results of up to 80% or more of schools are indistinguishable from the average school outcome so that real differences in school results can be only identified for a small minority of schools.

Using data on average test scores from 48 English junior schools, Goldstein [1997] found that that the uncertainty intervals indicated that the results of three-quarters of the schools were not statistically different from the average. A further study of average test scores for 76 primary schools in Hampshire found that the uncertainty intervals overlapped for 80 per cent of schools [Goldstein et.al. 2000]. A recent study has shown that in one selected local education authority in England the average scores of 50% of secondary schools were statistically indistinguishable from the overall average for all schools in the authority [Goldstein & Leckie 2008; see also Benton et.al. 2003].

Similar results have been obtained from US studies. The study of fourth grade reading and math results of elementary schools in North Carolina found that among schools near the national average in size (between 65 and 75 students with valid test scores), the margin of error (uncertainty interval) extended from approximately the 25th to the 75th percentile [Kane & Staiger 2002a]. That is, it wasn't possible to distinguish between the average results of 50% of schools.

A study of school value added results in 60 California schools concluded that it was only possible to reliably distinguish the top third of schools from the bottom third [Betebenner 2004]. This means it was not possible to distinguish the results of the bottom third of schools from the middle third, or those of the top third from the middle third.

In Australia, an ACER study found that 84 per cent of the uncertainty intervals for school reading scores in PISA 2000, adjusted for SES and gender, overlapped the average score of all schools [Rowe 2004]. As the author of this study states:

> ....it illustrates that attempts to separate or rank schools in the form of 'league tables' are subject to considerable uncertainty....Interpretation of estimates of individual schools is problematic, misleading and potentially irresponsible. Unfortunately, similar to their counterparts in the UK and the USA, Australian politicians and senior bureaucrats currently advocating the publication of such PI [performance indicators] 'league tables', *are naively 'stomping around' in an uninformed epistemopathological fog*. [13]

The uncertainty problems associated with comparisons of school results was emphasised in a recent study of the extent to which current school performance can be used as a guide to future performance, using the results of 266 secondary schools across England [Leckie & Goldstein 2009]. The study estimated school outcomes at the end of secondary school for the current intake cohort. Predicting future performance on the basis of current performance adds another layer of uncertainty into school comparisons. It found that the uncertainty intervals for the predicted value added estimates were 3.5 times as wide as those for the current value added estimates. As a result, it found that almost no schools are significantly different from the average school and very few schools can be predicted to be significantly different from each other. These results are likely to be even stronger for primary schools since these are, on average, much smaller than the average secondary school.

Professor Margaret Wu from the University of Melbourne has found that the NAPLAN results have large error margins [Wu 2009a]. The extent of the errors is quite large, even for individual students, and they are exacerbated at the class and school levels. Professor Wu found that measurement errors in annual 40-item tests, such as those being used in NAPLAN, would lead to about 16 per cent of students appearing to go backward when they had actually made a year's progress. She said this is a conservative estimate as it does not take account of other sources of error such as the assumption that two tests are assessing the same content. The errors could well be larger.

While the size of the measurement error reduces for classes and schools, they are still quite large. For example, Professor Wu found that the statistical uncertainty around the average results on these tests for classes of 30 students is equivalent to more than six month's learning. Many schools around Australia only have this many students or less participating in the NAPLAN tests. For schools, with two classes of 30 students tested the error could amount to about four months of learning.

A cursory inspection of the guide to statistical error posted on the My School website demonstrates that the margin of error is quite large. For example, the margin of error for 30 students participating in literacy and numeracy tests for Year 3 is of the order of 25-30 points where the national average score is around 400 while for 20 students it is around 30-35 points [see http://www.myschool.edu.au/TellMeMore.aspx]. This means that the score range for these schools is from about 385 to 415. A large number of schools reported average scores within this range and the error margin means that these schools cannot be statistically distinguished from each other.

Many parents would see interpret one school with a score of 415 as being significantly better than one with an average score of 385. Such an interpretation would be incorrect. It is not possible to make reliable comparisons or rankings of schools within the margins of error

because they may reflect chance differences in school performance rather than real differences. It means that comparisons of school results are mostly identifying lucky and unlucky schools, not good and bad schools.

Consequently, there is a strong likelihood that parents may be misled in choosing a school. Some schools may be recognised as outstanding while others are identified as unsuccessful simply as the result of chance and not because of actual programs and teaching practice. It also means that current school performance is highly misleading as a guide to future school performance.

Statistical errors in school results also make it difficult to identify effective school practices. It may mislead decision-makers and schools in recommending and adopting particular educational programs. Action taken to assist less successful schools may appear more effective than it is in practice.

6.2 School reporting should be statistically valid and reliable

Reporting the results of individual schools should be statistically valid and reliable so as not to mislead parents, schools and the public. The national education ministerial council agreement on the principles and protocols for reporting school results states that statistical error margins should be published with school results to ensure accurate interpretation [MCEETYA 2009]. It also states that the data used should be valid and reliable.

However, the My School website fails to comply with these requirements. My School only provides estimates of statistical error for various cohort sizes taking the tests. No information is provided on My School about the number of students who take the tests at each age level in each school. It is therefore impossible to apply the information provided about statistical error to any school. Nor is it possible to compare the results of different schools accurately because the margins of error for each test in each year level cannot be ascertained.

The federal and state education ministers should ensure that the national protocols for reporting school results are adhered to by the Australian Curriculum, Assessment and Reporting Authority [ACARA]. They should require ACARA to report the statistical margin of error for each test result reported in each school report on the *My School* website. The error margins should be reported for each test result to ensure accuracy, assist interpretation of results and ensure that parents are not misled.

These steps to better inform the public about the accuracy and reliability of school results have been recommended by expert statistical authorities such as the UK Statistics Commission [Statistics Commission 2004: para 19] and the Royal Statistical Society [Bird et.al. 2005: 16] as well as the National Center for Research on Evaluation, Standards and Student Testing in the United States [Linn 2001a: 31; see also Linn 2001b: 4; Baker & Linn 2002: 21, 24].

7. Conclusions

One of the main arguments in support of publishing school results is that it will help parents choose the right school for their children; help determine successful education practices; and help in deciding whether to take action on schools not succeeding. Information used for these purposes has to be quality information. Above all, it must be reliable and capable of making accurate distinctions between the results of different schools.

The school results published on My School and in league tables published in the media are an inaccurate and unreliable measure of school quality. They are apt to mislead for many reasons.

The results are significantly influenced by differences in the student composition of schools. Differences in school results are caused by differences in the socio-economic, ethnic and Indigenous composition of schools. They are also affected by differences in the proportion of students with disabilities participating in the tests.

This is well recognised by My School. It attempts to overcome this problem by publishing comparisons of so-called "like schools", or "statistically similar" schools. However, My School provides misleading and unreliable comparisons of test results of so-called 'like schools' because its measure of like schools is flawed [Cobbold 2010]. It does not consistently compare like with like. Indeed, it is systematically biased against government schools in comparison with private schools.

School results are also strongly influenced also by other external factors such as student absenteeism, student turnover, school size, school funding, parent involvement in learning at home, and the proportion of students receiving private tutoring. The results published on My School do not take account of these factors.

The results published on My School are also a selective measure of education. While of crucial importance, literacy and numeracy are far from the only aspect of student learning. Schools can do well in other important areas of learning such as supporting the personal and social development of students, arts and music and science, but not so well in literacy and numeracy. Failure to take account of the contribution schools make in other important areas of childrens' learning may give a distorted and inaccurate view of school quality.

School results may be artificially boosted by being manipulated in various ways. Overseas overseas experience shows that many schools resort to poaching high achieving students from other schools, denying entry to, or expelling, low achieving students, suspending low achieving students on test days, holding back students in grades not tested, increasing use of special dispensations for tests, encouraging students to take courses whose results are not used to compare schools and outright cheating. Extensive academic studies also show that test results are manipulated by schools in various ways to improve their ranking.

Several of these ways of manipulating school results are already being used in Australia. During the recent NAPLAN tests there were many instances of schools encouraging low achieving students to stay home during the tests, leaking of tests beforehand to alert teachers about questions and teachers helping students with answers and changing answers.

The overseas experience with publishing school results and league tables suggests that manipulation of school results is likely to increase in the future as a result of the pressure placed on teachers and principals to improve school results.

Published school results and league tables also mislead when measurement and sampling errors on school results are not reported. Many technical studies of school results and school league tables have demonstrated that chance differences account for a significant proportion of the differences in school test scores. The margin of error can be exceptionally large in measuring improvement which means that the results of the large majority of schools are

indistinguishable from each other. In addition, there can be considerable fluctuations in student achievement between years, especially in smaller schools.

This level of error wreaks havoc when comparing school results. It is not possible to make reliable comparisons or rankings of schools because they may reflect chance differences in school performance rather than real differences. Such comparisons are mostly identifying lucky and unlucky schools, not good and bad schools.

For all these reasons, parents choosing a school on the basis of published results may be misled. Some schools may be wrongly recognised as outstanding while others are identified as unsuccessful simply as the result of chance factors or the nature of the student intake. It may lead parents to choose a school of lesser quality than its results indicate.

These factors also make it difficult to identify effective school practices. Decision-makers and schools may be misled in recommending and adopting particular educational programs. Action taken to assist less successful schools may appear more effective than it is in practice.

The prospects that these problems can be overcome by changes to My School appear limited.

There is little likelihood of improvement in the accuracy of the measure of the socio-economic composition of schools. For example, it may be thought that the inherent bias against government schools in comparison with private schools caused by the area-based measure of socio-economic status used by My School could be overcome by resort to information obtained directly from families. However, a large proportion of families choose not to provide information about their income, education and occupation on enrolment forms and those not providing this information appear to be largely concentrated in the lower income, education and occupational groups. According to NAPLAN [2009], 17 to 25% of families of students at different Year levels do not provide parent education and occupation information on enrolment forms. The average literacy and numeracy results for students of these families (grouped as ‘non-stated’) are similar to those of students whose parents completed Year 12 and work in low skilled occupations. In addition, once children are enrolled it appears that schools make little effort to update information as family circumstances change.

Similar problems arise in obtaining more detailed information on school composition according to different ethnic sub-groups as not all families provide information on their background. Even today, some Indigenous families are reluctant to identify themselves as such.

There is also little prospect that the incentives for schools to manipulate and rort their results can be reduced without abandoning My School. Certainly, overseas systems that have been in place for up to 20 years have had little success in this regard. Manipulation and rorting continue to be a feature of systems that publish school results and league tables.

Manipulation and rorting of results is an inevitable outcome of publishing school results because of its “high stakes” implications for school reputations and the careers of teachers and principals. It is an example of a well known phenomenon in social science research called Campbell’s law, named after one of the most pre-eminent social scientists of the last century. Campbell’s law states:

> The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. [Wikipedia 2010]

When this law was first formulated, its author specifically applied it to education testing:

> Achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence**.** But when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways.

Campbell's law is a feature of "high stakes" performance reporting everywhere it is used. It seems that the designers of school performance reporting in Australia were oblivious to the extensive literature in economics and management theory which documents the inevitable corruption of quantitative indicators and the perverse consequences of performance incentives which rely on such indicators [for example, see Dranove et.al. 2002; Rothstein 2008].

The one area where some improvement could be made to My School to reduce the scope for misleading parents and the public is to report the margin of error on each school results for each subject tested at each Year level. However, this reporting would need to be accompanied by very strong caveats warning the public of the implications of the margins for error for comparing school results. There will be a strong temptation for parents to conclude that, say, a 30 point difference in results between two small schools means one is better than another, despite the fact that the two results are not statistically distinquishable.

References

Ainley, J.; Frigo T.; Marks, G. N.; McCormack, S.; McMillan J.; Meiers, M. & Zammit, S.A. 2000. The Measurement of Language Background, Culture and Ethnicity for the Reporting of Nationally Comparable Outcomes of Schooling. Report for the National Education Performance Monitoring Taskforce of the Ministerial Council for Education, Employment and Youth Affairs, Australian Council for Educational Research, August. Available at: http://www.mceecdya.edu.au/verve/_resources/languagebackground_file.pdf

Allen, R. & West, A. 2009. Religious Schools in London: School Admissions, Religious Composition and Selectivity. *Oxford Review of Education*, 35 (4): 471-494.

Arce-Ferrer, A.; Frisbie, D. A. & Kolen, M. J. 2002. Standard Errors of Proportions Used in Reporting Changes in School Performance With Achievement Levels. *Educational Assessment*, 8 (1): 59-75.

Auditor-General, Victoria 2009. *Literacy and Numeracy Achievement.* Victorian Government Printer, Melbourne.

Audit Office of NSW 2008. *Improving Literacy and Numeracy in NSW Public Schools : Department of Education and Training.* Audit Office of NSW, Sydney.

Baker, E. L. & Linn, R.t L. 2002. Validity Issues for Accountability Systems, CSE Technical Report 585, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, December. Available at: http://www.cse.ucla.edu/products/rsearch.asp

Benton, T.; Hutchison, D.; Schagen, I. & Scott, E. 2003. Study of the Performance of Maintained Secondary Schools in England. Report for the National Audit Office, National Foundation for Educational Research, November.

Berliner, D. 2009. Poverty and Potential: Out-of-School Factors and School Success. Education and the Public Interest Center & Education Policy Research, Arizona State University and the University of Colorado, March. Available at: http://epicpolicy.org/publication/poverty-and-potential

Betebenner, Damian 2004. An Analysis of School District Data Using Value-added Methodology. Report 622, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, March. Available at: http://www.cse.ucla.edu/products/summary.asp?report=622

Bird, S. M.; Cox, D.; Farewell, V. T.; Goldstein, H.; Holt, T. & Smith, P. C. 2005. Performance Indicators: Good, Bad and Ugly. *Journal of the Royal Statistical Society* A, 168 (1): 1-27.

Booher-Jennings, J. 2005 Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, 42 (2), 231–268.

Cobbold, T. 2009. The Free Market and the Social Divide in Education. *Dissent* 31: 12-19.

Cobbold, T. 2010. Like School Comparisons Do Not Measure Up**:** An Analysis of Flaws in Like School Comparisons on *My School*. Save Our Schools Research Paper, February. Available at: http://www.saveourschools.com.au

Cresswell, J. 2004. *Immigrant Status and Home Language Background: Implications for Australian Student Performance in PISA 2000*. Australian Council for Educational Research, Camberwell, October. Available at: http://www.acer.edu.au/documents/PISA_2000Immigrants.pdf

Cullen, J. & Reback, R. 2006. Tinkering Towards Accolades: School Gaming Under A Performance Accountability System. In T. Gronberg and D. Jansen (eds.), Improving School Accountability: Check-Ups or Choice, *Advances in Applied Microeconomics*, 14: 1-34.

Dranove, D.; Kessler, D.; McClellan, M. & Satterthwaite M. 2002. Is More Information Better: The Effects of 'Report Cards' on Health Care Providers. Working Paper 8697, National Bureau of Economic Research, Cambridge, MA, January. http://www.nber.org/papers/w8697

Dunn, B. 2009. Student Movement 2007 and 2008: NT Government Schools. School for Social and Policy Research, Charles Darwin University, July. Available at: http://www.cdu.edu.au/sspr/documents/DET_Student_Movement_2007_and_2008.pdf

Figlio, D. 2006. Testing, Crime and Punishment. *Journal of Public Economics*, 90: 837-851.

Figlio, D & Getzler, L 2006. Accountability, Ability and Disability: Gaming the System. In T. Gronberg & D. Jansen (eds.) Improving School Accountability: Check-Ups or Choice, *Advances in Applied Microeconomics*, 14: 35–49.

Gillard, J. 2008. Issues: COAG, National Education Agreement. Transcript, 1 December. Available at: http://www.deewr.gov.au/Ministers/Gillard/Media/Transcripts/Pages/Article_081202_073114.aspx

Gillard, J. 2009a. Speech to the National Public Education Forum. Old Parliament House, Canberra, 27 March. Available at: http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article_090327_173128.aspx

Gillard, J. 2009b. Interview with Alan Jones, 2GB, Transcript, 5 August. Available at: http://www.deewr.gov.au/Ministers/Gillard/Media/Transcripts/Pages/Article_090806_095222.aspx

Goldstein, H. 1997. Methods in School Effectiveness Research. *School Effectiveness and School Improvement*, 8 (4): 369-395.

Goldstein, H. & Leckie, G. 2008. School League Tables: What Can They Really Tell Us? *Significance*, June, 67-69. Available at: http://www.cmm.bristol.ac.uk/team/HG_Personal/Full%20Publications%20-%20download/Table%20of%20publications.htm

Goldstein, H.; Huiqi, P.; Rath, T. & Hill, N. 2000. The Use of Value Added Information in Judging School Performance. Institute of Education, University of London. Available at: http://www.cmm.bristol.ac.uk/team/HG_Personal/Full%20Publications%20-%20download/Table%20of%20publications.htm#1997

Haveman, R. & Wolfe, B. 1995. The Determinants of Children Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, 33 (4): 1829-1878.

Heilig, J. & Darling-Hammond, L. 2008. Accountability Texas-Style: The Progress and Learning of Urban Minority Students in a High-Stakes Testing Context. *Educational Evaluation and Policy Analysis*, 30 (2): 75-110.

Hill, A.; Navin, F. & Lynch, A. 2009. Coming to Grips with Student Mobility and Policy Implications: A Case Study from Regional Queensland. Paper presented to the Australian Association for Education Research Annual Conference, 29 November- 3 December, Canberra.

Jacob, B. A. 2005. Accountability, Incentives and Behaviour: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics* 89 (5-6): 761-796.

Jacob, B. A. & Levitt, S. 2003. Rotten Apples: An Investigation of the Prevalence and Predictions of Teacher Cheating. *Quarterly Journal of Economics*, 118 (3): 843-877.

Jennings, J. & Beveridge, A. 2009. How Does Test Exemption Affect Schools' and Students' Academic Performance? *Educational Evaluation and Policy Analysis*, 31 (2): 153-175.

Kane, T. J. & Staiger, D. O. 2001. Improving School Accountability Measures. Working Paper 8156, National Bureau of Economic Research, Cambridge, MA, March. Available at: http://www.nber.org/papers/w8156

Kane, T. J. & Staiger, D. O. 2002a. Volatility in School Test Scores: Implications for Test-Based Accountability Systems. In Diane Ravitch (ed.), *Brookings Papers in Education Policy 2002*, Brookings Institution Press, Washington DC: 235-283.

Kane, T. J. & Staiger, D. O. 2002b. The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives*, 16 (4): 91-114.

Kane, T. J.; Staiger, D. O. & Geppert, J. 2002. Randomly Accountable. *Education Next*, Spring.

Koretz, Daniel 2008. *Measuring Up: What Educational Testing Really Tells Us*. Harvard University Press, Cambridge, MA.

Leckie, G. & Goldstein, H. 2009. The Limitations of Using School League Tables to Inform School Choice. *Journal of the Royal Statistical Society* Series A, 172: 835-851.

Lemke, R.J; Hoerandner, C.M & McMahon, R.J. 2006. Student Assessments, Non-test Takers, and School Accountability. *Education Economics*, 14 (2): 235-250.

Linn, R. L. 2001a. The Design and Evaluation of Educational Assessment and Accountability Systems. CSE Technical Report 539, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, April. Available at: http://www.cse.ucla.edu/products/rsearch.asp

Linn, R. L. 2001b. Reporting School Quality in Standards-Based Accountability Systems. Policy Brief 3, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, Spring. Available at: http://www.cse.ucla.edu/products/policy.html

Linn, R. L. & Haug, C. 2002. Stability of School Building Accountability Scores and Gains. CSE Technical Report 561, Centre for the Study of Evaluation, National Centre for Research on Evaluation, Standards and Student Testing, University of California, Los Angeles. Available at: http://www.cse.ucla.edu/products/Reports/TR561.pdf

Marks, G. & McMillan, J. 2000. Social Background and Educational Outcomes: Preliminary Results from the Longitudinal Surveys of Australian Youth. Report for the National Education Performance Monitoring Taskforce, June.

Marks, G.; McMillan, J. & Hillman, K. 2001. Tertiary Entrance Performance: The Role of Student Background and School Factors. Longitudinal Surveys of Australian Youth Research Report No. 22, Australian Council of Educational Research, November. http://www.acer.edu.au/lsay/research.html

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) n.d. *National Report on Schooling in Australia 2008*. Curriculum Corporation, Melbourne. Available at: http://cms.curriculum.edu.au/anr2008/pdfs/anr2008_stats.pdf

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) 2009. Principles and Protocols for Reporting on Schooling in Australia, June. Available at: http://www.mceecdya.edu.au/mceecdya/nap_principles__protocols_for_rep_on_school_2009,27896.html

National Assessment Program for Literacy and Numeracy (NAPLAN) 2009. *National Report: Achievement in Reading, Writing, Language Conventions and Numeracy 2009*. Available at: http://www.naplan.edu.au/verve/_resources/NAPLAN_2009_National_Report.pdf

Nichols, S. & Berliner, D. 2007. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Harvard Education Press, Cambridge MA.

Noden, P. & West, A. 2009. Secondary School Admissions in England: Admission Forums, Local Authorities and Schools. Education Research Group, Department of Social Policy, London School of Economics and Political Science, December. Available at: http://www.risetrust.org.uk/forums.html

Radigan, J. 2007. The Accountability System's Back Door: Raising School Ratings by Losing Low Income and Minority Youth. Paper presented to the annual meeting of the American Educational Research Association, Chicago. Available at: http://centerforeducation.rice.edu/Research/Radigan/latest.htm .

Ravitch, D. 2009. The NCLB Paradox Enters a Twilight Zone. Bridging Differences blog, *Education Week*, 22 September. Available at: http://blogs.edweek.org/edweek/Bridging-Differences/2009/09/the_nclb_paradox_enters_the_tw.html

Reynolds, A. J.; Chen, C. & Hebers, J. E. 2009. School Mobility and Educational Success: A Research Synthesis and Evidence on Prevention. Paper presented at the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighbourhoods, Board on Children, Youth, and Families, National Research Council, 22 June, Washington DC. Available at: http://www.fcd-us.org/usr_doc/ReynoldsSchoolMobilityAndEducationalSuccess.pdf

Rothstein, R. 2008. Holding Accountability to Account: How Scholarship and Experience in Other Fields Inform Exploration of Performance Incentives in Education. National Center on Performance Incentives, Vanderbilt University, Nashville, February. Available at: http://www.performanceincentives.org/news__events/detail.asp?pageaction=ViewSinglePublic&LinkID=211&ModuleID=38&NEWSPID=1

Rowe, K. 2004. Analysing and Reporting Performance Indicator Data: 'Caress' the Data and User Beware! Paper presented at the Public Sector Performance and Reporting Conference, Sydney, April. Available at: http://www.acer.edu.au/documents/Rowe-IIR_Conf_2004_Paper.pdf

Rudd, K. 2008a. Quality Education: The Case for an Education Revolution in Our Schools. Address to the National Press Club, Canberra, 27 August. Available at: http://pmrudd.archive.dpmc.gov.au/node/5622

Rudd, K. 2008b. Interview with Mike Carlton and Sandy Aloisi. Radio 2UE, Transcript, 28 August. Available at: http://pmrudd.archive.dpmc.gov.au/node/5617

Rustique-Forrester, E. 2005. Accountability and the Pressures to Exclude: A Cautionary Tale from England. *Education Policy Analysis Archives*, 13 (26). Available at: http://epaa.asu.edu/ojs/article/viewFile/131/257

Simons, R.; Bampton, M.; Findlay, A. & Dempster, A. 2007. Student Mobility, Attendance, and Student Achievement: The Power of Implementing a Unique Student Identifier. Paper for the Annual Conference of the Australian Association for Research in Education. Available at: http://www.aare.edu.au/07pap/sim07568.pdf

Sims, D.P. 2008. Strategic Responses to School Accountability Measures: It's All in the Timing. *Economics of Education Review*, 27 (1): 58-68.

Sirin, S. R. 2005. Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75 (3): 417-453.

Skolverket (Swedish National Agency for Education) 2010. *What Influences Educational Achievement in Swedish Schools?* Stockholm. http://www.skolverket.se/sb/d/190

Sorin, R. & Iloste, R. 2006. Moving Schools: Antecedents, Impact on Students and Interventions. *Australian Journal of Education* 50 (3): 227-241.

Statistics Commission (UK) 2004. Value Added Measures in School Performance Tables. May. Available at: http://www.statscom.org.uk/C_144.aspx

Suliman, R. & McInerny, D. 2006. Motivational Goals and Student Achievement: Lebanese-background Students in South-Western Sydney. *Australian Journal of Education* 50 (3): 242-264.

Thomson, S. & De Bortoli, L. 2008a. *Exploring Scientific Literacy: How Australia Measures Up*. Australian Council for Educational Research, Camberwell.

Thomson, S. and De Bortoli, L. 2008b. *The Performance of Students in the Australian Capital Territory on PISA*. Australian Council for Educational Research, Camberwell.

Watson, L. 2008. Private Expectations and Public Schooling: The Growth of Private Tutoring in Australia. Paper Presented to the Australian Association for Research in Education National Conference, 30 November – 4 December, Brisbane. Available at: http://www.aare.edu.au/08pap/wat08692.pdf

West, A. & Pennell, H. 2000. Publishing School Examination Results in England: Incentives and Consequences. *Educational Studies*, 26 (4): 423-436.

West, A.; Hind, A. & Pennell, H. 2004. School Admissions and `Selection' in Comprehensive Schools: Policy and Practice. *Oxford Review of Education*, 30 (3): 347-369.

West, A.; Barham, E. & Hind, A. 2009. Secondary School Admissions in England: Policy and Practice. Report prepared for the Research and Information on State Education Trust, London. Available at: http://www.risetrust.org.uk/admissions.html

Wilson, D. 2004. Which Ranking? The Use of Alternative Performance Indicators in the English Secondary Education Market. *Public Money and Management*, 24 (1): 37-45.

Wikipedia 2010. Campbell's Law. Available at: http://en.wikipedia.org/wiki/Campbell%27s_Law

Wu, M. 2009a. Issues in Large-scale Assessments. Keynote Address Presented at Pacific Rim Objective Measurement Symposium, July 28-30, Hong Kong. Available at: http://www.promshk.org/

Wu, Margaret 2009b. NAPLAN for the Layperson. Educational Measurement Solutions, 22 October. Available at: http://www.edmeasurement.com.au/_docs/NAPLAN_ForLayPerson.pdf

Xia, N. 2010. Family Factors in Education. Dissertation, Pardee Rand Graduate School, RAND Corporation, Santa Monica. Available at: http://www.rand.org/pubs/rgs_dissertations/RGSD256/

Like School Comparisons Do Not Measure Up

An Analysis of Flaws in Like School Comparisons on *My School*

Submission to the Senate Education Committee Inquiry into the Administration and Reporting of NAPLAN

Part 4

SAVE OUR SCHOOLS

June 2010

http://www.saveourschools.com.au

Contents

Key Points

- *My School* makes misleading and unreliable comparisons of test results of so-called 'like schools' because its measure of like schools is flawed and omits many factors outside the control of schools which affect test results. It does not consistently compare like with like.

- The Index of Community Socio-Educational Advantage (ICSEA) which is used to measure the socio-economic status (SES) of schools is flawed. It attributes each student with the average SES of the area in which they live rather than the actual SES of their family. This leads to misclassifications of students because high and low income families often live in the same areas.

- The comparisons of 'like schools' systematically and unfairly favour private schools over government schools. The average SES of private schools is artificially lowered by ICSEA while the average SES of government schools is artificially inflated because high income families choose private schools at double the rate of low income families. This leads to comparisons of unlike schools rather than like schools.

- The 'like school' comparisons on *My School* are also misleading because they ignore other differences in the student composition of schools which strongly influence school test results, including differences by gender, ethnic sub-groups and students with disabilities.

- The 'like school' comparisons on *My School* do not take account of large differences in school funding between the states and between high SES private schools and high SES government schools.

- The 'like school' comparisons on *My School* do not take account of the impact of several other factors on school test results, including high student mobility between schools; school size differences, student selection and private tutoring.

- ICSEA is already 4 years out of date, being based on the 2006 Census data, and fails to take account of the changing social and economic circumstances of families and the areas in which they live, such as those arising from the global financial crisis.

- ICSEA may mismeasure the SES of some schools because it fails to distinguish between families with and without school-age children and because the family income and qualifications data it uses may be distorted by relatively high non-response rates to the Census questions.

- Comparisons of test results of 'like schools' can be distorted by schools manipulating and rorting their test results to artificially boost their rankings on school league tables.

- Because of its flaws and omissions, ICSEA exaggerates the differences in quality between 'like schools' and thereby misleads those who choose schools or make policy decisions based on these comparisons.

- An independent public review of ICSEA should be conducted to find a better way of determining like schools.

Summary

The *My School* website purports to enable comparisons of school results amongst socio-educationally similar schools - so-called 'like schools'. However, these comparisons are unreliable because the measure of like schools is flawed and it fails to take account of many factors outside the control of schools which affect test results. It does not consistently compare like with like.

There are four basic flaws in the so-called like school comparisons on *My School* which create much potential for schools to be wrongly and unfairly compared with schools that are significantly different:

- The Index of Community Socio-Educational Advantage used to determine like schools is an inaccurate measure of the socio-economic status (SES) of schools;
- Differences in school composition according to gender, ethnic sub-groups and students with disabilities, all of which influence school results, are ignored;
- Other factors outside the control of schools which influence school results, such as student mobility between schools and private tutoring of students, are ignored; and
- Differences between schools in funding, size and student selection practices also influence school results but are ignored.

ICSEA is an inaccurate measure of the socio-economic status of schools

ISCEA is an inaccurate measure of the school SES for several reasons, and therefore can lead to comparisons of unlike schools:

- It systematically over-estimates disadvantage in private schools and under-estimates disadvantage in government schools because high SES parents in low SES areas are more likely to enrol their children in private schools;
- It may under-estimate disadvantage in schools because many households fail to report their income and qualifications, and are not included in the data used to measure school SES;
- It does not distinguish between families with and without school-aged children;
- It does not fully measure family wealth;
- Schools with similar ICSEA ratings may have quite different SES profiles;
- The data used is out of date;
- The SES ratings are open to manipulation and fraud by schools falsifying student addresses.

The bias in the measurement of disadvantage in private and government schools is a major flaw in the comparisons of like schools on *My School*. It means that private schools are systematically favoured in comparisons with government schools because their average SES will be higher than as measured by ICSEA and the average SES of government schools with which they are compared will be lower than as measured by ICSEA. Private schools will then appear to have better average results than their 'like' government school counterparts because school results are strongly influenced by the SES composition of schools. Thus, 'like school' comparisons will be inaccurate and misleading.

This flaw in ICSEA occurs because each student is attributed with the average SES of the area in which they live rather than the actual SES of their family and because high SES families enrol their children in private schools at double the rate of low SES families.

The systematic bias in favour of private schools could be compounded if the relatively high non-response rates to Census questions on key ICSEA variables are concentrated among low SES families. In the 2006 Census, 13% of family households did not state, or only partially stated, their income while 18% of people in the 25-54 age groups did not state their non-school qualifications.

It is also a concern that ICSEA is already out of date because it is based on 2006 Census data and it will become even more out of date in the next few years. Census surveys are only carried out every five years and it takes up to two years for the SES data to become available after each Census. This means ICSEA will be seven years out of date by the time the data from the next Census becomes available. Significant social change can occur over such a period. For example, unemployment has increased significantly since the last Census due to the global financial crisis and this may affect the ICSEA ratings of schools.

Other key differences in student composition are ignored

While SES accounts for the large part of the influence of background factors on school performance, ICSEA ignores other background factors which also have a significant influence on school results. These include:

- Gender;
- Students from different ethnic backgrounds; and
- Students with disabilities.

The failure to take these differences in student composition into account is likely to invalidate many so-called ‘like school’ comparisons.

Girls consistently achieve higher literacy results than boys. This is likely to be very significant in comparing outcomes in all boy schools and all girl schools with similar SES profiles. On average, all boy schools will tend to have lower levels of average literacy achievement than all girl schools with a similar SES composition.

It is also likely that, on average, all girl schools will have higher outcomes than schools with a similar SES profile with more or less equal proportions of female and male students. The male population of mixed schools is likely to lower their average achievement compared to all girl schools.

ICSEA fails to take account of differences in the ethnic composition of schools. This has also has important implications for the accuracy and validity of ‘like school’ comparisons because there are significant differences in the average results of students from different ethnic backgrounds. For example, the average results of Chinese students are well above those of Middle Eastern and Pacific Islander students. Schools with a similar ICSEA rating could have quite different average results simply because some have a high proportion of students of Chinese origin while others have a high proportion of Middle Eastern or Pacific Islander students.

ICSEA also ignores differences in the proportion of students with disabilities in so-called ‘like schools’. Schools with higher proportions of students with disabilities participating in tests may have lower average results than other schools with a similar ISCEA value. Government schools could be disadvantaged by ‘like school’ comparisons because they have higher proportions of students with disabilities than private schools.

Other factors outside the control of schools are ignored

The 'like school' comparisons also ignore other factors outside the control of schools that influence school results, including differences in:

- Student mobility between schools; and
- Private tutoring of students.

Some schools with a similar ICSEA rating may have lower results because they have a high proportion of students who often change school. Students who move school often tend to have lower average results than other students.

Moreover, many families today engage private tutors for their children. Differences between schools in the proportion of families who resort to private tutoring may be reflected in differences in school results between so-called 'like schools'.

Other differences between schools are ignored

The 'like school' comparisons on *My School* also ignore other key differences between schools that are not related to teaching and curriculum, but which also influence school results. These include differences in:

- Funding;
- School size; and
- Student enrolment procedures such as the selection of students.

My School fails to report school funding. Schools with similar ISCEA values may have vastly different levels of funding which may contribute to differences in school results. For example, in 2007-08 recurrent expenditure per government secondary school student in Western Australia was 43% higher than in Victoria while expenditure per primary school student in the ACT was 48% higher than in Victoria. In addition, total funding per student in many high SES private schools is up to double or more that of high SES government schools.

My School does not take account of differences in schools size which can affect the comparison of school results. Very large schools tend to have lower average results than small to medium sized schools. Small schools are much more likely to report large changes in average results from one year to the next because their results can be heavily influenced by the results of only 4 or 5 students.

The 'like school' comparisons on *My School* ignore student selection by many private schools and some government schools. Some schools may achieve higher results than others with a similar ICSEA rating because they can select higher achieving students and exclude lower achieving students.

In addition to all these flaws, comparison of test results of 'like schools' can also be distorted by schools manipulating and rorting their test results to artificially boost their rankings on school league tables.

ICSEA misleads about differences in school quality

Because of its flaws and omissions, ICSEA exaggerates the differences in quality between 'like schools' and thereby misleads those who choose schools or make policy decisions on the basis of these comparisons. For example, the variables incorporated in ICSEA explain about 70% of the variation in aggregated primary school results, leaving about 30% implicitly

attributed to differences in teaching, curriculum, pastoral care and other features of schools. If factors such as gender, ethnic sub-groups, students with disabilities, student mobility, school funding, school size and student selection practices were included in ICSEA its explanatory power could increase to 85-90% of the variation in school results. This would leave only 10-15% of the variation in school results as explained by differences in school quality.

Many reports, including some commissioned by the national council of ministers, have acknowledged a number of these flaws as issues in developing valid measures of school performance. In particular, they have noted the problems in using area-based measures of SES and the need to include gender and ethnic composition of schools in measuring school performance, but which are omitted in the *My School* comparisons. Yet, this advice has been ignored. An especially puzzling feature of the failure to take this advice is that one of the co-authors of several reports is a member of the expert panel advising on the development of ICSEA.

ICSEA in its current format is entirely inadequate for determining like schools. It needs a major overhaul. The systematic bias in favour of private schools generated by the use of an area-based SES measure can only be resolved by developing a consistent measure of individual family SES. ICSEA should also be revised to include gender, ethnic sub-groups, students with disabilities, student mobility, school funding and student selection practices.

A better way of determining like schools should be investigated by an independent public review.

An Analysis of the Flaws in Like School Comparisons on *My School*

1. Like school comparisons on *My School*

The My School website operated by the Australian Curriculum Assessment and Reporting Authority (ACARA) publishes literacy and numeracy results for similar schools (so-called 'like schools') across Australia. These are schools whose students have a similar social and economic background. The remoteness of schools and the proportion of Indigenous students is also taken into account in determining like schools. Each school is compared to up to 60 statistically socio-educationally similar schools (30 above and 30 below) as measured by their rating score on an Index of Community Socio-Educational Advantage (ICSEA) set with a mean value of 1000 [ACARA 2010a; 2010b; 2010c].

ICSEA is constructed from data on 14 key socio-economic factors that correlate with educational outcomes, as indicated by the National Assessment Program for Literacy and Numeracy (NAPLAN). The factors include 2 family income variables; 5 family education variables; 3 occupation variables, 1 unemployment variable and 3 other variables. The data for these variables are obtained from the 2006 Census of Population and Housing for census collection districts, areas consisting of about 225 households.

ACARA estimates that 68% of the variance in aggregated primary school outcomes is explained by the 14 socio-economic variables used to construct ICSEA together with the measure of school remoteness and the percentage of Aboriginal enrolments [ACARA 2010a]. It estimates that 59% of the variance in aggregated secondary school outcomes is explained by these factors. The implication is that the remaining 30 or 40% of the variance in school results is due to school quality factors.

The ICSEA rating for each school is obtained by matching the addresses of students at each school with its census collection district and attributing the socio-economic status (SES) characteristics of that area to each individual student. The data from each area from which the school population is drawn is aggregated up to the school level to obtain the average for each SES variable at the school level. The consolidated school level SES data is fed into the ICSEA formula to calculate the school ICSEA rating.

These ratings are used to construct groups of similar schools for each individual school, comprising 30 schools above and below each school's ICSEA rating; that is, there is a different like school group for each school. *My School* provides a link from each school's report card to the results for each of these 60 'like schools' so that their results can be compared.

There are four basic flaws in these so-called like school comparisons which create much potential for schools to be wrongly and unfairly compared with schools that are significantly different:

- ICSEA is an inaccurate measure of the SES of schools;

- Differences in school composition according to gender, ethnic sub-groups and students with disabilities, all of which influence school results, are ignored;
- Other factors outside the control of schools which influence school results, such as student mobility between schools and private tutoring of students, are ignored; and
- Differences between schools in funding, size and student selection practices also influence school results but are ignored.

2. ICSEA is an inaccurate measure of the socio-economic status of schools

ICSEA is an inaccurate measure of the school SES for several reasons, and thus can lead to comparisons of unlike schools:

- It systematically over-estimates disadvantage in private schools and under-estimates disadvantage in government schools because higher SES parents tend to enrol their children in private schools;
- It may under-estimate disadvantage in government schools because many households fail to report their income and qualifications, and are not included in the data used to measure school SES;
- It does not distinguish between families with and without school-aged children;
- It does not include a measure of family wealth;
- Schools with similar ICSEA ratings may have quite different SES profiles;
- The data used is out of date;
- The ratings are open to manipulation and fraud by schools falsifying student addresses.

2.1 ICSEA over-estimates disadvantage in private schools and under-estimates disadvantage in government schools

ICSEA systematically over-estimates the level of socio-economic disadvantage in private schools and under-estimates disadvantage in government schools. Consequently, *My School* compares the test results of supposedly similar private and government schools, but which may have large differences in the SES composition of their enrolments.

This occurs because ICSEA is derived from area-based Census data and not data on individual families attending each school. In effect, each student is attributed with the average SES of the area in which she/he lives rather than the actual SES of each student's family. This creates the potential for students to be misclassified to a higher or lower SES than their family.

Assigning a value of socioeconomic status to a student on the basis of the area in which they live will introduce a potential error and the magnitude of the error will be greater when the social background of those living in the area is relatively heterogeneous. [Ainley & Long 1995: 53; see also Marks et.al.2000]

In some areas, the population may be clustered around the average SES score, while in others the distribution may be more diverse, that is, with significant proportions of both lower and higher SES households.

A relatively disadvantaged area is likely to have a high proportion of relatively disadvantaged people. However, such an area is also likely to contain people who are not disadvantaged, as well as people who are relatively advantaged. When area level indexes are used as proxy measures of individual level socio-economic status, many people are likely to be misclassified. This is known as the ecological fallacy. [ABS 2008b: 3]

Studies in Australia show that this effect exists. A study by the Australian Council of Educational Research (ACER) has shown that the correlation between individual and Census collection district measures of SES for a national sample of secondary school students was unacceptably low [Ainley & Long 1995].

The Australian Bureau of Statistics (ABS) has demonstrated that some highly advantaged families live in low SES areas and some disadvantaged families live in high SES areas. In an analysis of Census data for Western Australia it found that individual and family relative socio-economic disadvantage was quite diverse within small areas [Baker & Adhikari 2007]. About 20 per cent of people in the most disadvantaged quartile of the individual SES measure lived in Census collection districts that were in the highest three deciles of the area-based Index of Relative Socio-economic Disadvantage (IRSD). Over a third of people in the bottom quartile lived in areas in the top five IRSD deciles and six per cent of people in the lowest group in the individual based SES measure lived in collection districts found in the highest IRSD decile.

On the other hand, nearly 20 per cent of people in the most advantaged quartile for individual SES lived in areas that were classified in the bottom three deciles of the IRSD. Over a third of people in the most advantage quartile lived in areas in the bottom five deciles. Five per cent of people in the highest individual based SES group lived in the collection districts found in the lowest IRSD decile.

> The analysis shows that there is a large amount of heterogeneity in the socio-economic status of individuals and families within small areas. These findings indicate that there is a high risk of the ecological fallacy when SEIFA is used as a proxy for the socio-economic status of smaller groups within an area and there is considerable potential for misclassification error. [Baker & Adhikari 2007: 1]

The potential for error in measuring the SES of schools by using area-based Census data is quite significant because high SES families enrol their children in private schools at double the rate of low SES families. In 2006, 47% of high income families across Australia enrolled their children in private schools compared to 24% of low income families [Preston 2007]. Fifty-five per cent of high income families enrolled their children in private secondary schools compared to 26% of low income families. It is likely that the rate at which high income families in low SES areas enrol their children in private schools is even higher than the average rate for Australia.

As a result, ICSEA artificially lowers the average SES of private schools and artificially raises the average SES of government schools. Private schools that attract high SES students from low SES areas will receive a lower ICSEA rating, thus over-estimating their actual level of disadvantage. The government schools serving these low SES areas will receive a rating which takes account of high SES families in the area. However, because children of these high SES families are not actually enrolled in the government schools their ICSEA value will be higher than their actual SES. The level of disadvantage in government schools is thus under-estimated by ICSEA.

This systematic bias in the mismeasurement of the SES of government and private schools is not just a feature of low SES areas. It is repeated across all areas containing both low and high SES families. Therefore, government and private schools measured as similar by ICSEA are likely to be dissimilar.

This bias in the measurement of disadvantage in private and government schools has major implications for the comparison of 'like school' results. It will tend to systematically favour private schools which will appear to have higher average results than their 'like' government school counterparts because school results are strongly influenced by the SES composition of schools. Thus, the 'like school' comparisons will tend to be inaccurate and misleading. ACARA has acknowledged this limitation of ICSEA, but its only response was to make secret adjustments to the like school comparisons where evidence was available that the area SES values do not accurately reflect the actual student composition of a school. Such adjustments are only likely to take account of the more obvious "Geelong Grammar" phenomenon whereby high SES farming families in low SES rural areas who enrol their children in wealthy private schools carrying their low SES area rating with them. Mismatches between area-based SES and actual school SES are likely to be much more extensive than such obvious examples because high SES families are more likely to choose private schools.

2.2 ICSEA is likely biased because many low income families do not reveal their income and qualifications

ICSEA is also likely to be biased because not all people return Census forms or fully answer Census questions and the response rates vary for different questions. The non-response rates for the factors used to construct ICSEA vary considerably. In particular, 13% of family households did not state their income or only partially stated their income while 18% of people in the 25-54 age groups did not state their non-school qualification and 8% of this age group did not state their highest level of school education [ABS 2007]. In addition, 7% of occupied private dwellings did not respond to the internet access question. In comparison, only 2% of people in the 24-54 age groups did not state their occupation and 5% did not state their language and proficiency in English.

Household income, non-school qualifications and the level of school education all have a significant influence on student achievement and the relatively high rate of missing data in the Census for these factors for could be a source of bias in the ICSEA ratings. For example, if it is assumed that those who did not state their income, non-school qualification and highest level of school education are all from low SES households, the proportion of these families nationally and in different regions will be under-estimated. Similarly, if it is assumed that non-responding dwellings do not have Internet access, there could be a corresponding upward bias in the estimated connectivity levels of about 7% at the national level. ACARA does not appear to have investigated the potential implications of this missing data for the ICSEA ratings of schools and comparisons of like schools.

There does not appear to be any direct evidence on the characteristics of those who failed to fully or partially answer these questions. However, there is some evidence to suggest that non-responses to surveys requesting information on income and education qualifications tend to be concentrated amongst low SES families. For example, there is a high non-response rate to questions on family income and occupations on school enrolment forms. According to NAPLAN [2009], 17 to 25% of families of students at different Year levels do not provide parent education and occupation information. The average literacy and numeracy results for students of these families (grouped as 'non-stated') are similar to those of students whose parents completed Year 12 and work in low skilled occupations.

This evidence suggests that the non-response rates for the Census data used to construct ICSEA could be largely due to non-responses by lower SES families. The absence of this data may have several implications. It may mean that the overall strength of the relationship

between low SES and student outcomes is under-estimated as is the variance in aggregated school outcomes explained by the variables used to construct ICSEA. It may also mean that the number of families with low income, education and qualifications in low SES areas is underestimated so that the average SES of these areas is also under-estimated, the latter being the weighted average of all the families in the area, including higher SES families, who responded to the Census questions.

In this event, the ICSEA ratings of government and private schools serving these low SES communities would be over-estimated, that is, the level of disadvantage in these schools is under-estimated. This would not affect like school comparisons if the non-response rates were similar across all low SES Census collection districts. However, this is unlikely to be the case. Significantly different response rates between low SES areas could result in some schools being given a higher rating than warranted and being wrongly compared to schools with a higher ICSEA rating.

If the non-responses to the Census questions on income and non-school qualifications are concentrated among low SES families, it could also compound the systematic bias in ICSEA which favours private schools in comparison with government schools because higher SES families are more likely to choose private schools.

It is possible that the missing data does not have any significant impact on the ICSEA ratings. However, ACARA does not appear to have done any sensitivity analysis on the ratings to establish their robustness in the presence of missing Census data. It is incumbent on ACARA conduct and report on such an analysis. The ICSEA ratings of schools have significant implications for the comparison of school results, the reputations of schools and funding by the Commonwealth Government.

2.3 ICSEA fails to distinguish between families with and without school-age children

ICSEA may also provide inaccurate measures of school SES because the version used to determine like school comparisons does not distinguish between families with and without school-aged students. This may have implications for the measurement of school SES in areas where there significant differences in the SES of families with and without school-age children. It may lead to some schools being defined as low SES because of high concentrations of pensioners and young unemployed in the area, while families with school-age children are well-off. In such cases, these schools would be incorrectly classified as low SES schools and wrongly compared to actual low SES schools when they should be compared with higher SES schools.

My School has acknowledged this problem, but chose to ignore it even though it increases the extent of the variance in school outcomes explained by the ISCEA variables.

2.4 ISCEA does not fully measure family wealth

A recent discussion paper published by the Commonwealth Department of Education notes that a number of studies have found that there is a high correlation between family wealth measures and educational participation and attainment [DEEWR 2009]. Wealth appears to exert a notable influence on education participation and achievement over and above that due to parent education and occupation.

Typically, income is used as a surrogate for wealth because data on family wealth is not available from the Census. However, while family income and wealth will often overlap as measures of SES, income is not fully adequate as a measure of the resources available to families. Low income does not necessarily consistently indicate low SES as some families can be income poor and asset rich. For example, some families who own businesses in industries such as agriculture, mining, fishing, finance investment and tourism often experience widely fluctuating incomes over different periods and may have low incomes for several years, during which they are able to draw on their assets to support the education of their children in various ways.

If the period of low income coincides with the Census which *My School* draws on to construct its measure of SES some schools may be incorrectly assessed as low SES despite many of their families having considerable wealth.

2.5 Schools with similar ICSEA ratings may have quite different SES profiles

Schools having the same ICSEA rating may have significantly different student profiles which can lead to differences in average school results of so-called like schools.

Each school's ICSEA rating is a weighted average of the SES of the resident Census collection districts of its students. As such, it does not reveal differences in the proportion of students from areas with different SES scores. For example, one school with an ICSEA value of, say, the scaled mean of 1000, could have all its individual student addresses clustered around the mean while another with a similar score could have a large proportion of students resident in low SES addresses and a large proportion whose addresses are in high SES areas. It is possible that the average test results of the former could be significantly higher than the latter because of the latter's higher proportion of low SES students. Comparing the test scores of such schools could therefore be misleading.

2.6 ISCEA is already out of date

ISCEA is already out of date by nearly four years as it is based on 2006 Census data and it will become even more out of date before new Census data is available in 2013 or 2014.

Since the last Census, major changes have occurred that have potential to cause the actual SES of many schools to differ from their ISCEA value. For example, the global financial crisis has increased unemployment in Australia and the pattern of unemployment varies across different regions. This may mean that the actual SES of some schools in these regions is now significantly below their measured SES according to the 2006 Census data and below that of schools considered by *My School* to be statistically similar.

Ongoing changes in the demographic profiles of suburbs between Census dates may also change the actual SES of schools in some areas and invalidate the 'like school' comparisons determined by ISCEA values. A classic example from the past was the large scale movement of higher income families into inner suburban areas to replace low SES families. This is still going on in some cities. Such changes have the potential to cause the actual SES composition of some inner city schools to be significantly different from that indicated by their ISCEA value.

Social change is an ongoing phenomenon. Census data are collected every five years and it usually takes 2 or 3 years to construct the new SES index values. The next Census is in 2011

and new data will not become available for at least two years after that. It means that the *My School* measure of SES could become outdated by 7 or 8 years. Significant social change affecting school SES can occur within this period.

2.7 ISCEA ratings of schools are open to manipulation

ISCEA ratings of private schools are open to manipulation by submitting false student addresses to ensure a lower rating and favourable test score comparisons with lower SES schools. The SES funding model for private schools creates similar incentives for schools to submit false student addresses to ensure a lower school SES measure and therefore greater funding.

My School obtains residential addresses for private school students from the Commonwealth Department of Education. They are required to be submitted to the Department for the purposes of SES funding by the Commonwealth Government. A report last year by the National Audit Office found that there was scope for the Department to improve its procedures to check the accuracy of student enrolment data for private schools [ANAO 2009].

Addresses for government school students are supplied by state and territory education departments. These too could be open to manipulation as departments do not conduct close audits of information submitted by schools.

This may appear to be a somewhat fanciful or unrealistic possibility at this stage. However, overseas experience with comparisons of school results shows that schools are under extreme pressure to maintain or improve their ranking against other schools and often resort to underhand ways to manipulate their results to look better. Lax security and audit procedures create scope for fraudulent practices.

3. Like school comparisons ignore other differences in school composition

The *My School* comparisons of the test results of 'like schools' ignore other differences in the student composition of schools which have a significant effect on school results. While SES accounts for the large part of the influence of background factors on school performance, there are also other background factors which, if not taken into account, could invalidate like-school comparisons. These include differences in school composition by gender, ethnic sub-groups and students with disabilities.

Several studies, including studies co-authored by a member of the ICSEA expert advisory panel, have found these factors to be significant in explaining the variation in student achievement. It is strange that ACARA has not included these factors in the construction of ICSEA as it would increase the explanatory power of the index.

3.1 *My School* ignores differences in the gender composition of schools

Surprisingly, ICSEA ignores the impact of gender on school results. Many studies in Australia have highlighted significant differences in educational outcomes for females and males.

It is well recognised that females demonstrate significant and consistently higher levels of average literacy achievement than males [Marks et.al. 2001; Rothman 2002; Rothman &

McMillan 2003]. There seems to be a consistent pattern across age groups [Cresswell et.al. 2002]. At the primary level, males have significantly lower levels of achievement in literacy than females. At the secondary school level, males are more likely than females to leave school before completing Year 12, and average scores on end-of-school assessments are lower for males than for females.

These findings are confirmed by recent results. The latest NAPLAN results show that in each literacy domain at each Year level, the percentage of students at or above the national minimum standard is greater for females than for males [NAPLAN 2009]. Average results for females significantly exceed those for males. Similarly, the latest PISA results show significantly higher literacy outcomes for by 15 year old females than for males [Thomson & de Bortoli 2008]. Other information also indicates that the Year 12 results for government and private all female schools are significantly higher than those of all male schools in the same regions [Parker 2008].

On the other hand, there is evidence that males slightly outperform females in numeracy, although the evidence is mixed for different age levels [Rothman 2002; Rothman & McMillan 2003]. The most recent NAPLAN results show that average scores for male students are generally slightly higher than for females. However, there is little difference in the proportion of females and males at or above the minimum numeracy standards. The PISA results show slightly higher average achievement for males than for females [Thomson & de Bortoli 2008].

These differences in gender achievement are unlikely to be significant for like school comparisons involving schools with similar proportions of female and male students. However, they are likely to be very significant in comparing outcomes between all male schools and all female schools with similar SES profiles. On average, all male schools will tend to have lower levels of average literacy achievement than all female schools with similar SES composition. It is also likely that, on average, all female schools will have higher outcomes than schools with a similar SES profile but which have more or less equal proportions of female and male students. That is, the male population of these schools is likely to lower their average achievement compared to all female schools. Similarly, all male schools may have slightly higher levels of numeracy achievement than all female schools.

Thus, ICSEA should take account of significant variations in the proportion of female enrolments between schools in order to ensure a better measure of 'like schools'. Inclusion of a gender variable would increase the explanatory power of the index.

3.2 *My School* ignores differences in the ethnic composition of schools

My School does not distinguish the ethnic profile of schools. Children from particular ethnic groups are often concentrated in particular areas and schools, and children from some ethnic communities are highly concentrated in government schools. Performance disparities between 'like' schools may reflect differences in ethnic composition rather than differences in school practices.

The ethnic background of students is excluded from ICSEA because the correlations between the percentage of people who do not speak English well and student achievement were below statistical significance [ACARA 2010]. This is not surprising because when all students from ethnic backgrounds, or from non-English speaking backgrounds, are considered as a group they achieve at similar levels to students of Australian-born parents [Cresswell 2004;

Cobbold 2009]. There is little evidence that students with language backgrounds other than English have poorer educational outcomes. This has been the case for many years [Sturman 1997; Ainley et.al. 2000].

However, there are significant differences in student achievement between different groups of students from non-English speaking backgrounds. ACARA has mistakenly analysed only a very broad category of people and ignored differences within the group of people who do not speak English well. Asian students, in particular, Chinese students, appear to achieve at higher levels than other groups. Students from some ethnic groups achieve much lower average results.

There is evidence that the average results of Lebanese students and Pacific Islander students are well below those of Chinese students. A study of Year 10 Certificate results for English, Mathematics and Science in three high schools in south-western Sydney found substantial differences between the results of students from different ethnic backgrounds [Suliman & McInerny 2006]. It found a much higher percentage of Lebanese students were achieving in the lower grades in all subject areas than Chinese and Vietnamese students. A much lower percentage of Lebanese students achieved the top grades than Chinese and Vietnamese students. The average results of Lebanese students were significantly below those of Chinese and Vietnamese students.

An analysis of data from the Longitudinal Surveys of Australian Youth shows that achievement in literacy and numeracy in Year 9 varies widely among students from language backgrounds other than English [Marks & McMillan 2000]. When ethnic background was measured by nine categories relating to father's country of birth, some ethnic groups showed higher Year 9 achievement levels than those students with fathers born in Australia, while students from other groups showed lower Year 9 achievement levels. This finding was confirmed by further analysis of university entrance scores. It found that students classified as Asian performed substantially better than students whose fathers were born in Australia when adjusted for socio-economic background [Marks et.al. 2001]. On the other hand, the mean scores of students with Middle Eastern, North African and Pacific Islander ancestries were significantly below those of Asian students.

A study carried out by ACER for the National Education Performance Monitoring Taskforce of the Ministerial Council for Education, Employment and Youth Affairs shows that average numeracy test scores for Year 9 students from the Middle East, North Africa and the Pacific Islands are significantly below those of Asian students [Ainley et.al. 2000].

A later ACER report which analysed the 2000 PISA results for Australian immigrant children found that average reading and mathematics scores for Chinese and other Asian language students were significantly higher than for students with Middle Eastern home languages [Cresswell 2004]. In terms of proficiency levels, nearly 50% of Chinese students and 29% of Other Asian students achieved at the top two reading levels compared to 15% of Middle Eastern students. In contrast, only 9% of Chinese students did not achieve expected minimum standards compared to 24% of Middle Eastern students.

Such disparities in outcomes for students from different ethnic backgrounds are concealed by aggregate measures as used by ACARA. The ACER report to the Performance Monitoring Task Force stated:

....it needs to be acknowledged that there is a great diversity in the educational outcomes of students from ethnic minorities. Students from some ethnic or language backgrounds perform better than others and the use of aggregated data conceals these differences. Consequently, it is important to collect detailed data on the cultural and language backgrounds of students. [6]

It recommended that when using country of birth and language measures to report educational outcomes, students should be grouped into several categories [Ainley et.al. 2000: 32]. It is curious that this recommendation was not adopted by ACARA when one of the co-authors of the report is on the expert advisory panel for ICSEA.

A recent report by ACER to the MCEETYA Expert Working Group on reporting for school evaluation and accountability also recommended that national reporting on school performance should include the language backgrounds of students in schools should be taken into consideration in any evaluation of school performance [ACER 2008: 24].

The failure of ICSEA to take into account the different ethnic composition of schools has important implications for the accuracy and validity of like school comparisons. Performance disparities between so-called like schools may reflect differences in ethnic composition rather than differences in school practices. For example, schools with a similar ICSEA ranking could have a high proportion of students of Asian origin while others have a high proportion of Middle Eastern or Pacific Islander students which accounts for the difference in average results. Parents and others using the like school comparisons of My School could therefore be misled about the comparative quality of teaching and curriculum in these schools.

3.3 *My School* ignores students with disabilities

While there is provision for students with disabilities to be exempted from national tests this is not automatic and they are encouraged to participate by all Australian governments. These students receive special assistance for the tests. The results of schools serving these students are reported in the same way on *My School* as for other schools [ACARA 2010b].

Many schools have higher proportions of students with disabilities than others. Schools with higher proportions of students with disabilities participating in tests may have lower average results than other schools with a similar ISCEA value. A report by ACER to the MCEETYA Expert Working Group on school evaluation and accountability stated:

...in schools with large percentages of students with special needs, this variable may need to be taken into account in understanding average levels of school attainment....Care also is required in comparing outcomes for schools with significant proportions of special educational needs students. [Masters et.al. 2008: 25]

In addition to the bias inherent in the ISCEA ratings, government schools could be disadvantaged by ‘like school’ comparisons because they have a higher proportion of students with disabilities than private schools. For example, the proportion of students with disabilities in total government school enrolments in NSW and Victoria is over 5% compared to about 1% in Anglican schools.

4. *My School* ignores the influence of other outside factors on school results

4.1 *My School* ignores differences in student mobility between schools

The 'like school' comparisons on *My School* ignore the impact of differences in student turnover in schools. Some schools may have lower results because they have a high proportion of students who often change school.

Student mobility in Australia can be high. For example, a Queensland study has shown that 16% of all primary school students moved school 2 or 3 times in five years and 4% moved school 4 or more times [Simons et.al. 2007]. Another Queensland study also shows very high student mobility in some schools [Hill et.al. 2009]. Other data shows that nearly 40 per cent of students in the Northern Territory change schools in any one year, apart from those who go on to a higher stage of schooling [Dunn 2009]. Even in a high income city such as Canberra, there are schools where annual turnover is over 30% at times.

While student mobility is higher amongst low SES families than for high SES families, it is also significant for high SES families. For example, the Queensland study found that nearly 30% of low SES primary school students and 13% of high SES students changed school 2 or more times in five years.

Many studies show that students who move school often tend to have lower average results than students who remain at the same school; although there is also evidence that it has little effect [Sorin & Iloste 2006; Simons et.al. 2007]. A recent syntheses of research studies on the effects of school mobility on achievement and dropout rates found that the large majority of studies conclude that children who move school 3 or more times have significantly lower reading and mathematics achievement and are more often to drop out of school [Reynolds et.al. 2009].

Thus, schools with similar ISCEA values could have large differences in average literacy and numeracy test scores because of differences in the proportion of students who often change schools. It also begs the question of the extent to which NAPLAN results can be credited to a school when a significant proportion of its students are only recent enrolments.

4.2 *My School* ignores the impact of private tutoring

The use of private tutoring by families is so extensive that it has been called a 'shadow education system' [Watson 2008]. The use of private tutoring by school-aged children in Australia is increasing and expenditure on private tutoring is increasing as a proportion of total household expenditure on children's education [Watson 2008]. Private tutoring tends to be used more by higher income families. The wealthiest households (those in the top 20 per cent of the income distribution) spend double the amount of an average household on private tutoring.

Differences in the extent to which families resort to private tutoring outside school is another factor affecting 'like school' comparisons. Schools with the same ISCEA value may have different proportions of students engaged in private tutoring. This may be a factor influencing comparisons of government and private schools with the same ISCEA value. Higher SES families can better afford private tutoring for their children A private school may achieve

higher results than a government school simply because a larger proportion of its families use private tutoring.

In addition, changes in school results from one year to the next may be influenced by changes in the proportion of families who use private tutoring. If a higher proportion of families engage private tutoring in any one year a school will receive a boost to its measured performance compared to another 'like school' even though there was no change in teaching effectiveness during the year.

5. Like school comparisons ignore many other different features of schools

Schools differ in many ways apart from their socio-economic composition and the influence of other outside school factors. Some other key ways in which schools differ and which may influence school results are ignored by My School in providing comparisons of test results between so-called 'like schools'. These include differences in funding, size and the selection of students.

5.1 *My School* ignores vast funding differences between schools

My School fails to report school funding. Schools with similar ISCEA values may have vastly different levels of funding which may contribute to differences in school results.

There are very large differences in recurrent expenditure on government schools in different states. For example, in 2007-08 recurrent expenditure per government secondary school student in Western Australia was 43% higher than in Victoria while expenditure per primary school student in the ACT was 48% higher than in Victoria [Productivity Commission 2010]. In addition, total funding per student in many high SES private schools is up to double or more that of high SES government schools [Cobbold 2010].

Comparing the results of schools with similar ISCEA values while ignoring such vast expenditure differences between schools is not comparing like with like. It is comparing unlike schools.

The Federal Minister of Education has long promised that school funding levels would be reported on My School. However, the Government reneged on this promise. It is now only promised at a future date and may not come to fruition because of strong opposition from private school organisations.

5.2 *My School* ignores the impact of differences in school size

My School comparisons of 'like schools' ignore the impact of school size. It allows comparisons of large and small low SES schools and comparisons of large and small high SES schools.

Many research studies suggest that smaller primary and middle schools have higher levels of student achievement than very large schools while the findings for high schools are mixed [Garrett et.al. 2004; Howley 2002a; Hicks & Rusalkina 2004; McMillen 2004; Stevenson 2006]. These findings confirm earlier reviews of the research literature [Fowler & Walberg 1991].

There is also an extensive research literature which shows that students from low SES backgrounds achieve better results in smaller schools [Cobbold 2006]. Small schools with high concentrations of students from low SES backgrounds tend to have higher average results than large schools with similar concentrations.

Thus, some differences in results between schools with the same ICSEA ratings may be due to size differences or to the interaction of school size with socio-economic composition.

In addition, the results of small schools may significantly affected by the results of only a few students. Studies show that small schools are much more likely to report large changes in average results from one year to the next, both positive and negative [Kane & Staiger 2001; Linn & Haug 2002; Wu 2009]. School results can be heavily influenced by the results of 4 or 5 students in schools with small numbers of students in each year level and by the transfer of students between schools. Such volatility can lead to misleading and unfair comparisons between schools of different size but with the same ISCEA value.

5.3 *My School* ignores student selection by private schools

The 'like school' comparisons on *My School* ignore differences in enrolment practices by government and private schools which can affect comparisons of test results between schools with the same ICSEA value.

Private schools generally exercise control over which students they enrol. There is no requirement on private schools to accept all comers as is the case with government schools, apart from a small number of selective government schools. Private schools can also set their fee levels in a way that serves to exclude some students.

Private schools can use their control over enrolments to maximise their school results by selecting higher achieving students and excluding lower achieving students, an option not generally available to government schools. Thus, private and government schools with similar ICSEA values could have significantly different results because of differences in their ability to select their students.

6. Like school comparisons are affected by manipulation and rorting of school results

Extensive academic studies show that test results are manipulated and rorted by schools in various ways to improve their ranking [Cullen & Reback 2006; Figlio 2006; Figlio & Getzler 2002; Jacob 2005; Nicholls & Berliner 2007; Reback 2008].

There is evidence of this already happening in Australia at the senior secondary level. For example, many private schools make excessive use of special dispensations for Year 12 exams [Patty 2008; 2009a; 2009b; Tomazin 2009]. Some West Australian schools are pushing Year 12 students to choose easier subjects so that they can avoid exams that are used to rank school performance [Hiatt 2009]. Some Sydney private schools have forced some of their Year 11 students to change schools or do their HSC at TAFE because they may not achieve high enough results [Patty 2009c; Thompson 2009].

Following the overseas experience, it can be expected that Australia will soon see a variety of practices to manipulate school results. Practices commonly used in England and the United States include poaching high achieving students from other schools, denying entry to, or

expelling, low achieving students, suspending low achieving students on test days, increasing use of special dispensations for tests, encouraging students to take courses whose results are not used to compare schools and outright cheating.

Differences in the extent of such manipulation of school results will invalidate comparisons of test scores between schools with similar ISCEA values.

7. Conclusions

There are an extensive number, indeed, a litany, of flaws in the like school comparisons on *My School*. As a result, *My School* does not consistently compare 'like' with 'like'. A number of examples where comparisons between supposedly 'like schools' involve comparisons between very unlike schools have already come to light in newspaper reports around Australia.

Some of the flaws in the *My School* comparisons are more significant than others. A major flaw is the use of an area-based measure of SES which could substantially distort comparisons of schools, especially between government and private schools because higher SES families are more likely to choose private schools than low SES families.

Another major flaw is the exclusion of differences in the gender and ethnic composition of schools, factors which have been shown to have a significant effect on school results. Similarly, schools with higher proportions of students with disabilities participating in the national tests are likely to be unfairly compared with similarly rated schools with lower proportions of these students.

The failure to take account of student mobility, especially for small schools, large differences between schools in funding and size and student selection by some schools could also generate comparisons between very unlike schools.

The omission of these and other variables affecting school outcomes is quite significant for the comparison of like schools. It means that a substantial part of the variation in school results is incorrectly attributed to differences under the control of schools, for example, better teachers, curriculum, pastoral care, etc., instead of to omitted variables.

The technical analysis carried out by ACARA shows that much of the variation between schools can be explained by the factors incorporated in the ICSEA. According to ACARA, the socio-economic and other variables included in ICSEA explain nearly 70% of the variation in aggregated primary school outcomes. The remaining 30% of the variation is unexplained and implicitly attributable to differences in the quality of 'like schools'.

If gender, ethnic sub-groups, students with disabilities, school funding, school size, student mobility, school size and student selection were all properly taken into account, the index could well lift its explanatory power to 85-90% of the differences in school results. This would leave about 10-15% of the variation in school results to be explained by quality and other features of schools instead of the 30% currently implied by ICSEA. That is, ICSEA incorrectly attributes about half or more of the unexplained 30% to differences in school quality when it should be attributed to other factors outside the control of schools.

In other words, ICSEA exaggerates the extent of differences in quality between ‘like schools’ and misleads those who choose schools or make policy decisions on the basis of these comparisons.

It is notable that many reports, including some commissioned by the national council of ministers, have acknowledged a number of these problems as issues in developing accurate measures of school performance. In particular, they have noted the problems in using area-based measures of SES and the need to include gender and ethnic composition of schools in measuring school performance, but which are omitted in the My School comparisons. Yet, this advice has been ignored. An especially puzzling feature of the failure to take this advice is that one of the co-authors of several reports is a member of the expert panel advising on the development of ICSEA.

It is clear then that ICSEA in its current format is entirely inadequate for determining like schools. It needs a major overhaul. The systematic bias in favour of private schools generated by the use of an area-based SES measure can only be resolved by developing a consistent measure of individual family SES. ICSEA should also be revised to include gender, ethnic sub-groups, students with disabilities, student mobility, school funding, school size and student selection practices.

An independent public inquiry should be established to review the flaws in ICSEA and the failure of *My School* to consistently compare like with like.

References

Adhikari, Pramod 2006. Socio-Economic Indexes for Areas: Introduction, Use and Future Directions. Research Paper, Australian Bureau of Statistics. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/PrimaryMainFeatures/1351.0.55.015?OpenDocument

Ainley, John and Long, Michael 1995. Measuring Student Socioeconomic Status. In John Ainley; Brian Graetz; Michael Long and Margaret Batten, *Socioeconomic Status and School Education*, Australian Government Publishing Service, Canberra, June.

Ainley, John; Frigo Tracey; Marks, Gary N.; McCormack, Silvia; McMillan Julie; Meiers, Marion and Zammit Susan A. 2000. The Measurement of Language Background, Culture and Ethnicity for the Reporting of Nationally Comparable Outcomes of Schooling. Report for the National Education Performance Monitoring Taskforce of the Ministerial Council for Education, Employment and Youth Affairs, Australian Council for Educational Research, August. Available at: http://www.mceecdya.edu.au/verve/_resources/languagebackground_file.pdf

Australian Bureau of Statistics (ABS) 2007. Census of Population and Housing. Available at: http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Census+data

Australian Bureau of Statistics 2008a. An Introduction to Socio-Economic Indexes for Areas (SEIFA). Information Paper. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/PrimaryMainFeatures/2039.0?OpenDocument

Australian Bureau of Statistics 2008b. Socio-Economic Indexes for Areas (SEIFA) - Technical Paper. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/PrimaryMainFeatures/2039.0.55.001?OpenDocument

Australian Curriculum, Assessment and Reporting Authority 2010a. Index of Community Socio-Educational Advantage (ICSEA), *My School* Technical Paper. Available at: http://www.myschool.edu.au/Resources/pdf/My%20School%20ICSEA%20TECHNICAL%20PAPER%2020091020.pdf

Australian Curriculum, Assessment and Reporting Authority 2010b. About ICSEA. *My School* Fact Sheet. Available at: http://www.myschool.edu.au/Resources/pdf/My%20School%20FACT%20SHEET%20ABOUT%20ICSEA%2020100120.pdf

Australian Curriculum, Assessment and Reporting Authority 2010c. Frequently Asked Questions. Available at: http://www.myschool.edu.au/Resources/pdf/My%20School%20FAQs%2020100120.pdf

Australian National Audit Office 2009. *Funding for Non-Government Schools*. Audit Report No.45, 2008–09 Performance Audit, Canberra. Available at: http://www.anao.gov.au/uploads/documents/2008-09_Audit_Report_45.pdf

Baker, Joanne and Adhikari, Pramod 2007. Socio-economic Indexes for Individuals and Families. Research Paper, Analytic Services Branch, Australian Bureau of Statistics, June. Available at: http://www.abs.gov.au/AUSSTATS/abs@.nsf/ProductsbyCatalogue/4F209039E8F94F0CCA25733F0013802B?OpenDocument

Cobbold, Trevor 2006. The Interaction of School Size and Socio-Economic Status on Student Performance. Save Our Schools, October. Available at: http://www.saveourschools.com.au/

Cobbold, Trevor 2009. The Free Market and the Social Divide in Education. *Dissent* 31: 12-19.

Cobbold, Trevor 2010. Government Funding for Elite Private Schools: An Upper Class Welfare Policy. Save Our Schools, forthcoming.

Cresswell, John 2004. *Immigrant Status and Home Language Background: Implications for Australian Student Performance in PISA 2000*. Australian Council for Educational Research, Camberwell, October. Available at: http://www.acer.edu.au/documents/PISA_2000Immigrants.pdf

Cresswell, John; Rowe, Ken and Withers, Graeme 2002. *Boys in School and Society*. Australian Council for Educational Research, Camberwell. Available at: http://www.acer.edu.au/documents/BOYSEDU_BoysInSchoolAndSociety.pdf

Cullen, Julie B. and Reback, Randall 2006. Tinkering Toward Accolades: School Gaming Under a Performance Accountability System. *Advances in Applied Microeconomics* 14: 1-34.

Department of Education, Employment and Workplace Relations (DEEWR) 2009. Measuring the Socio-economic Status of Higher Education Students, Discussion Paper, December. http://www.deewr.gov.au/HigherEducation/Documents/LowSES_Discussionpaper.pdf

Dunn, Bruce 2009. Student Movement 2007 and 2008: NT Government Schools. School for Social and Policy Research, Charles Darwin University, July. Available at: http://www.cdu.edu.au/sspr/documents/DET_Student_Movement_2007_and_2008.pdf

Figlio, David N. 2006. Testing, Crime and Punishment. *Journal of Public Economics*, 90 (4-5): 837-851.

Figlio, David N. and Getzler, Lawrence S. 2002. Accountability, Ability and Disability: Gaming the System, Working Paper 9307, National Bureau of Economic Research, Cambridge, October. Available at: http://www.nber.org/papers/w9307 .

Fowler, William J. and Walberg, Herbert J. 1991. School Size, Characteristics and Outcomes. *Educational Evaluation and Policy Analysis*, 13 (2), 189-202.

Garrett, Zoe; Newman, Mark and Elbourne, Diana 2004. *Secondary School Size: A Systematic Review*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, October. Available at: http://eppi.ioe.ac.uk/cms/

Hiatt, Bethany 2009. Schools 'Dumb Down' Year 12s to Lift Ranking. *The West Australian*, 13 July.

Hicks, Michael J. and Rusalkina, Viktoriya 2004. School Consolidation and Educational Performance: An Economic Analysis of West Virginia High Schools. Center for Business and Economic Research, Marshall University, May. Available at: http://www.marshall.edu/cber/research/index.htm.

Hill, Angela; Navin, Fiona and Lynch, Andrea 2009. Coming to Grips with Student Mobility and Policy Implications: A Case Study from Regional Queensland. Paper presented to the Australian Association for Education Research Annual Conference, 29 November- 3 December, Canberra.

Howley, Craig 2002. Small Schools. In Alex Molnar (ed.) *School Reform Proposals: The Research Evidence*. Education Policy Studies Laboratory, College of Education, Arizona State University, January.

Jacob, Brian 2005. Accountability, Incentives and Behaviour: Evidence from School Reform in Chicago. *Journal of Public Economics*. 89 (5-6): 761-796.

Kane, Thomas J. and Staiger, Douglas O. 2001. Improving School Accountability Measures. Working Paper 8156, National Bureau of Economic Research, Cambridge, March.

Linn, Robert L. and Haug, Carolyn 2002. Stability of School Building Accountability Scores and Gains. CSE Technical Report 561, Centre for the Study of Evaluation, National Centre for Research on Evaluation, Standards and Student Testing, University of California, Los Angeles. Available at: http://www.cse.ucla.edu/products/Reports/TR561.pdf

Marks, Gary and McMillan, Julie 2000. Social Background and Educational Outcomes: Preliminary Results from the Longitudinal Surveys of Australian Youth. Report for the National Education Performance Monitoring Taskforce, June.

Marks, Gary; McMillan, Julie; Jones, Frank L. and Ainley, John 2000. The Measurement of Socioeconomic Status for the Reporting of Nationally Comparable Outcomes of Schooling. Report to the National Education Performance Monitoring Taskforce, MCEETYA, March. Available at: http://www.mceecdya.edu.au/mceecdya/measure_socioeconomic_status,11969.html

Marks, Gary; McMillan, Julie and Hillman, Kylie 2001. Tertiary Entrance Performance: The Role of Student Background and School Factors. Longitudinal Surveys of Australian Youth Research Report No. 22, Australian Council of Educational Research, November. http://www.acer.edu.au/lsay/research.html

Masters, Geoff; Rowley, Glenn; Ainley, John and Khoo, Siek Toon 2008. Reporting and Comparing School Performances. Australian Council of Educational Research, December. Available at: http://www.curriculum.edu.au/verve/_resources/ACER_Report_on_Reporting_and_comparing_school_performances.pdf

McMillen, Bradley J. 2004. School Size, Achievement and Achievement Gaps. *Education Policy Analysis Archives*, 12 (58). Available at: http://epaa.asu.edu/epaa/v12n58/.

National Assessment Program for Literacy and Numeracy (NAPLAN) 2009. *National Report: Achievement in Reading, Writing, Language Conventions and Numeracy 2009*. Available at: http://www.naplan.edu.au/verve/_resources/NAPLAN_2009_National_Report.pdf

Nichols, Sharon L. and Berliner, David C. 2007. *Collateral Damage: How High Stakes Testing Corrupts America's Schools*. Harvard Education Press, Cambridge, Mass.

Parker, Maralyn 2008. Top Schools in Region 2008 – Albury to Inner West. *The Daily Telegraph*, 23 December. Available at: http://blogs.news.com.au/dailytelegraph/maralynparker/index.php/dailytelegraph/comments/top_schools_in_region_a_to_i/

Patty, Anna 2008. Private Schools Accused of Rorting the HSC. *The Sydney Morning Herald*, 29 December.

Patty, Anna 2009a. No 'Special' Haven at Selective Schools. *The Sydney Morning Herald*, 11 March.

Patty, Anna 2009b. HSC Help Most Often for Private Students. *The Sydney Morning Herald*, 18 December.

Patty, Anna 2009c. Schools Play Truant to Avoid Bad Marks. *The Sydney Morning Herald*, 20 October.

Preston, Barbara 2007. The Social Make-up of Schools. A report prepared for the Australian Education Union. Available at: http://www.aeufederal.org.au/Publications/Bprestonrep2007.pdf

Productivity Commission 2010. *Report on Government Services 2010*. School Education Attachment Tables, Table 4A.8. Available at: http://www.pc.gov.au/gsp/reports/rogs/2010

Reback, Randall 2008. Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics*, 92 (5-6): 1394-1415.

Reynolds, Arthur J.; Chen, Chih-Chih and Hebers, Janette E. 2009. School Mobility and Educational Success: A Research Synthesis and Evidence on Prevention. Paper presented at the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighbourhoods, Board on Children, Youth, and Families, National Research Council, 22 June, 2009, Washington DC. Available at: http://www.fcd-us.org/usr_doc/ReynoldsSchoolMobilityAndEducationalSuccess.pdf

Rothman, Sheldon 2002. Achievement in Literacy and Numeracy by Australian 14 year-olds, 1975-1998. Longitudinal Surveys of Australian Youth Research Report No 29, Australian Council of Educational Research, November. http://www.acer.edu.au/lsay/research.html

Rothman, Sheldon and McMillan, Julie 2003. Influences on Achievement in Literacy and Numeracy. Longitudinal Surveys of Australian Youth Research Report No 36, Australian Council of Educational Research, October. Available at: http://www.acer.edu.au/lsay/research.html

Simons, Roland; Bampton, Margo; Findlay, Andrea and Dempster, Alistair 2007. Student Mobility, Attendance, and Student Achievement: The Power of Implementing a Unique Student Identifier. Paper for the Annual Conference of the Australian Association for Research in Education. Available at: http://www.aare.edu.au/07pap/sim07568.pdf

Sorin, Reesa and Iloste, Rosemary 2006. Moving Schools: Antecedents, Impact on Students and Interventions. *Australian Journal of Education* 50 (3): 227-241.

Stevenson, Kenneth R. 2006. School Size and its Relationship to Student Outcomes and School Climate. National Clearinghouse for Educational Facilities, Washington DC, April. Available at: http://www.edfacilities.org/pubs/index.cfm.

Sturman, Andrew 1997. *Social Justice in Education*. Australian Education Review No. 40, Australian Council of Educational Research, Camberwell.

Suliman, Rosemary and McInerny, Dennis 2006. Motivational Goals and Student Achievement: Lebanese-background Students in South-Western Sydney. *Australian Journal of Education* 50 (3): 242-264.

Thomson, Sue and de Bortoli, Lisa 2008. *Exploring Scientific Literacy: How Australia Measures Up*. Australian Council for Educational Research, Camberwell.

Thompson, Judy 2008. Letter to the Editor. *The Sydney Morning Herald*, 20 December.

Tomazin, Farrah 2009. Private Students Get More Exam Aid. *The Age*, 12 February.

Watson, Louise 2008. Private Expectations and Public Schooling: The Growth of Private Tutoring in Australia. Paper Presented to the Australian Association for Research in Education National Conference, 30 November – 4 December, Brisbane. Available at: http://www.aare.edu.au/08pap/wat08692.pdf

Wu, Margaret 2009. NAPLAN for the Layperson. Educational Measurement Solutions, 22 October. Available at: http://www.edmeasurement.com.au/_docs/NAPLAN_ForLayPerson.pdf

Conclusions and Recommendations

Submission to the Senate Education Committee Inquiry into the Administration and Reporting of NAPLAN

Part 5

SAVE OUR SCHOOLS

June 2010

http://www.saveourschools.com.au

Table of Contents

1. Introduction

The former Federal Minister for Education claims that publishing school results for NAPLAN on the My School website will create pressures for school improvement and better inform parental choice of school. Parts 1 to 3 of this submission demonstrate that these claims are not substantiated by the evidence. Instead, the evidence suggests exercise of school choice, facilitated by the publication of such results often leads to a situation in which children, whose parents have the resources to move to another school, abandon a school serving a largely disadvantaged community, therefore further marginalising the school. Put simply, school choice advantages the already advantaged, and disadvantages the disadvantaged. It therefore does not contribute to the equity goals articulated by the now Prime Minister.

The former Education Minister also claims that publishing school results is needed to identify struggling schools. This submission argues firstly that there is ample information on disadvantage in schools, certainly enough to inform educational funding decisions, and that the reason that disadvantage remains such a major factor in educational outcomes in Australia is that governments, both State and Commonwealth, have failed to provide the funding to support the complex and often expensive programs need to change things. This submission also argues that any possible benefits from NAPLAN do not depend on publication of school results.

Finally, the former Education Minister has argued that publication of NAPLAN results is necessary simply because parents and the public have a right to be informed about school results. This submission does not reject the general principle of the right of the public, and in this case parents in particular, to information, but suggests where release of information is likely to enhance rather reduce disadvantage, there is a clear public benefit case for restricting the use of this information to areas where it might be beneficial.

2. The case against publishing school results and league tables

2.1 Publication of school results and league tables does not improve student achievement

As demonstrated in Part 1 of this submission, there is no conclusive research evidence that publishing school results improves school performance. Even the chief executive of ACARA, Peter Hill, admits there is little evidence to support his Minister's claim.

Very few studies have separately assessed the impact of reporting school results, but the key studies show no significant effect on student achievement. For example, a Brookings Institution study found "no discernable effect". A frequently cited study by Professor Eric Hanushek and Margaret Raymond of Stanford University on the positive effects of accountability measures found that reporting school results alone has no impact on student achievement.

Most studies assess reporting school results along with a range of 'high-stakes accountability' measures. These include rewards for improved performance and sanctions such as reconstitution of schools, replacing the principal or teachers, permitting students to enrol elsewhere, grade promotion standards and having a high school graduation exam.

The major studies of so-called 'high-stakes accountability' for school results show no significant gains in student achievement or reduced achievement gaps. A recent sophisticated

meta-analysis of 14 major accountability studies in the US found mixed effects on reading and mathematics achievement. Seven studies favoured states with high stakes testing and reporting, six studies had mixed or insignificant findings, and one study favoured states with low-stakes testing. Student accountability requirements, such as grade promotion standards, end-of-course and graduation exams, appear to contribute a major part of the positive effect shown in some 'high stakes accountability' studies.

The major academic reviews of the research studies on 'high stakes accountability' measures conclude that the evidence is mixed and that there is little scientific foundation for these policies.

There is also no conclusive evidence that reporting school results will increase student achievement by promoting choice and competition. The weight of evidence from the most comprehensive studies of the impact of choice and competition is that it does not improve student achievement once social and demographic factors are taken into account. As Professor of Economics at the University of Chicago and the co-author of *Freakonomics*, Steven Levitt, says of school choice, "the theory sounds great, but evidence confirming it has been hard to find" [Levitt 2007]. A recent study by the London School of Economics concluded that "choice and competition does not seem to be generally effective in raising standards".

Increased competition as a result of school comparisons and rankings appears to more lead to standardization rather than innovation in classroom practice and curriculum. A study published by the OECD summarising evidence on innovations in more market-driven education systems in over 20 countries found that competition between schools tends to promote uniformity rather than innovation and an emphasis on marketing rather than educational change [Lubienski 2009].

2.2 Publishing school results and league tables harm education

Part 2 of this submission demonstrated that, far from improving student achievement, there is strong evidence that publishing school results and league tables harm education because they:

- Narrow the curriculum;
- Distort teaching practice;
- Disadvantage low and high achieving students;
- Unfairly stigmatise low achieving students;
- Make it more difficult for low performing schools to retain high quality teachers;
- Discourage co-operation and collaboration between schools and teachers; and
- Increase social segregation and inequity in education.

The curriculum is narrowed in at least two ways. More time is devoted to the tested subjects of literacy and maths while untested subjects such as science, history, social studies, languages, arts and music, physical education and health receive much less time. Even recess gets cut. Second, in the subjects that are tested, greater emphasis is given to the areas that are most conducive to testing by multiple-choice questions and there is less teaching of more complex skills.

Teaching is distorted because schools and teachers tend to respond to pressure created by publication of school results and league tables by focusing more on teaching test-taking skills and practicing for tests. Publishing school results and league tables can undermine efforts to

improve the quality of teaching because they turn the notion of a “good teacher” into one who increases test scores.

Schools tend to concentrate on improving the results of students who are on the border of accepted benchmarks at the expense of both high and low achieving students. Publication of the results of individual schools and public rankings of schools may unfairly stigmatise and humiliate some students and alienate them from schooling. Published school results and league tables may be used as a job guide and make it more difficult for schools with low results to retain and attract quality teachers and principals. Publication of school results and competition for league table rankings can reduce collaboration between schools and between teachers within schools.

Publication of school results and league tables tends to increase socio-economic and ethnic segregation between schools which exacerbates inequity in education. Student learning needs in some schools increase without proportionate increases in resources to meet those needs and increasing concentrations of students from low socio-economic status families in some schools tend to lead to lower overall outcomes.

There is evidence of some of these effects already in Australia after only one year of publishing school results.

2.3 Published school results are misleading and unreliable

In addition to their harmful effects, school NAPLAN results are often an inaccurate and misleading meausure of school quality. Part 3 of this submission showed that published school results and league tables are likely to mislead parents who use them to inform their choice of school because:

- Differences in school composition affect school results;
- Many other factors outside schools influence school results;
- They are a selective measure of education;
- They are subject to manipulation and rorting; and,
- There may be significant statistical errors on school test results.

School results often more reflect the socio-economic background of its students. They are strongly influenced also by other external factors such as student absenteeism, student turnover, parent involvement in learning at home, and the proportion of students receiving private tutoring.

School results may be artificially boosted by being manipulated and rorted under the pressure to maintain rankings. Overseas studies show that schools often resort to poaching high achieving students from other schools; denying entry to, or expelling, low achieving students; suspending low achieving students on test days; using special dispensations for tests; and outright cheating. Some of this is already happening in Australia.

Differences in school test results may be largely due to chance because of inevitable measurement and sampling error in tests. Several studies, including one Australian study, show that the results of up to 80% of schools are statistically indistinguishable when measurement error is taken into account. Real differences in school results can be only identified for a small minority of schools.

In addition, the much greater variation in test scores within schools than between schools in Australia means that parents cannot be assured that their child will be in a "better class" in a high performing school than in a low performing school. The much greater variation in test results within schools means that any given school with higher scores than another is likely to have classes that perform worse than the best classes in the lower achieving school.

The combined effect of incomplete information on important socio-demographic variables, statistical reliability and rorting make the NAPLAN results quite unreliable for fine discriminations between schools, when most schools are statistically indistinguishable. It means that parents who make decisions about school choice based on what are statistically indistinguishable results may be misled. Perceived differences in school results which in fact are not statistically distinguishable may also mislead decision-makers and schools in recommending and adopting particular educational programs. Some education practices and programs could be incorrectly identified as having more success than others.

3. Like school comparisons are no answer

It is argued that so-called "like school", or "statistically similar" school, comparisons provide contextual information that obviates the harm of simplistic comparisons of school results and league tables.

Part 4 of this submission demonstrated that My School makes misleading and unreliable comparisons of so-called "like schools" because its measure of like schools is flawed and omits many factors outside the control of schools which affect test results. It does not consistently compare like with like.

The Index of Community Socio-Educational Advantage (ICSEA) which is used to measure the socio-economic status (SES) of schools is flawed. It attributes each student with the average SES of the area in which they live rather than the actual SES of their family. This leads to misclassifications of students because high and low income families often live in the same areas.

A major flaw is that comparisons of "like schools" on My School systematically and unfairly favour private schools over government schools. The average SES of private schools is artificially lowered by ICSEA while the average SES of government schools is artificially inflated because high income families choose private schools at double the rate of low income families. This leads to comparisons of unlike schools rather than like schools.

The "like school" comparisons on My School are also misleading because they ignore other differences in the student composition of schools which strongly influence school test results, including differences by gender, ethnic sub-groups and students with disabilities. They do not take account of large differences in school funding between the states and between high SES private schools and high SES government schools. Nor do they take account of the impact of several other factors on school test results, including high student mobility between schools; school size differences, student selection and private tutoring.

The measure of school SES is already 4 years out of date, being based on the 2006 Census data and will become up to 7 years out of date before the next Census data is available. ICSEA may mismeasure the SES of some schools because it fails to distinguish between families with and without school-age children and because the family income and

qualifications data it uses may be distorted by relatively high non-response rates to the Census questions.

Because of its flaws and omissions, ICSEA exaggerates the differences in quality between 'like schools' and thereby misleads those who choose schools or make policy decisions based on these comparisons.

"Like school" comparisons are, in effect, partial league tables and they incur the same problems as full league tables. They do not fully account for the influence of outside factors on school results. They narrow student learning in the same way. They provide the same incentives for schools to rig their results. The lowest ranked schools in each like school group will be pilloried and humiliated. They discourage collaboration and co-operation between schools. Moreover, the error margins on school NAPLAN results are likely to be so large as to invalidate many comparisons of like school results.

4. Revisions to My School will not resolve fundamental problems

A number of revisions are planned for My School in 2010 and 2011 [ACARA 2010]. They include measures to minimise misuse of My School data, providing additional contextual information for each school, enhancements of ICSEA and additional student outcome data. None of the measures will resolve the fundamental problems with My School outlined above. Some are likely to compound existing problems.

It is proposed that ACARA will investigate ways of deterring or preventing electronic collection of data from the My School website as a way of deterring the publication of league tables. It remains to be seen whether this is possible, but even if it is it only addresses the issue of league tables in the media.

The fundamental issue is publishing the results of individual schools. Using published school results to compare school performance and as a guide to school choice harms education and misleads parents and the public about school problems. This is exacerbated by the publication of league tables in the media.

My School itself will continue to publish what, in effect, are partial league tables in the form of "like school" comparisons and local area school comparisons. These have the same problems as full league tables and compound the problems of publishing school results.

The additional contextual information on schools will include funding, students with disabilities, students with a language background other than English, teacher level of expertise, and the results of parent, teacher and student satisfaction surveys. While these enhancements provide additional information with which to interpret schools' results, they will not reduce the harm of publishing school results.

Education Ministers have requested ACARA to investigate the feasibility and appropriateness of making use of student-level SES data on their parents' or carers' education and occupation. ACARA is also to investigate:

- obtaining updated and comprehensive home address data for all students to improve the accuracy of ICSEA in cases where area-based data are used;
- including within the ICSEA formula a variable to take account of the effect of language background other than English;

- improving the process for quality assuring ICSEA values for individual schools and, for those for which area-based data are used, identifying instances where the initial estimate is inappropriate.

Presumably, these measures are intended as a way to reduce the problems of using an area-based measure of school SES. Given the complexity of some of the variables and the difficulty in obtaining reliable disaggregated data, it remains to be seen whether these changes can enhance the accuracy of "like school" comparisons.

In principle, the inherent bias against government schools in comparison with private schools caused by the area-based measure of socio-economic status used by My School could be overcome by resort to information obtained directly from families. However, a large proportion of families choose not to provide information about their income, education and occupation on enrolment forms and those not providing this information appear to be largely concentrated in the lower income, education and occupational groups. According to NAPLAN [2009], 17 to 25% of families of students at different Year levels did not provide parent education and occupation information on enrolment forms. The average literacy and numeracy results for students of these families (grouped as 'non-stated') are similar to those of students whose parents completed Year 12 and work in low skilled occupations. In addition, once children are enrolled it appears that schools make little effort to update information as family circumstances change.

Similar problems arise in obtaining more detailed information on school composition according to different ethnic sub-groups as not all families provide information on their background. Even today, some Indigenous families are appareantly reluctant to identify themselves.

The incorporation of growth data on literacy and numeracy results for schools and the publication of senior secondary outcomes will add to the problems of My School in the harm it does to education and in misleading parents and the public about school quality. From a basic statistical perspective, the errors in estimates at the school level are already large. If change over time is compared, then the statistical errors of the two estimates add to make these estimates even more unreliable. It is therefore unlikely that examining changes over time will overcome the fundamental problem that most schools are statistically indistinguishable.

While the proposed revisions to My School may tighten up the analysis of value-added compared to expectation, they seem to be based on the idea that fine discriminations in school performance are what is required. This ignores the basic statistical problem of drawing conclusions about school quality when most schools are statistically indistinguishable. The proposed changes are unlikely to add much to the precision of the data.

The proposed revisions do not address the fundamental problem associated with publishing school results – the harm it does to education. The incentives for schools to narrow the curriculum and teaching will remain. They will remain for as long as school results continue to be published, even if the publication of league tables is somehow reduced or prohibited.

There is also little prospect that the incentives for schools to manipulate and rort their results can be reduced without abandoning My School. Certainly, overseas systems that have been in place for up to 20 years have had little success in this regard. Manipulation and rorting

continue to be a feature of systems that publish school results and league tables. Manipulation and rorting of results is an inevitable outcome of publishing school results because of its "high stakes" implications for school reputations and the careers of teachers and principals.

5. Other arguments to publish school results are not valid

As Education Minister, Julia Gillard made much of the proposition that reporting of school results is necessary to identify struggling schools and those in need of intervention programs and additional resources. This is a grave case of ignorance. It is nonsense that publishing school results is needed to better allocate resources. This information was already available to education departments and schools. Governments have failed to provide the necessary resources and support for struggling schools

A further argument used to justify publication of school results is that parents and the public have a right to this information. This is also a flawed case.

There can be no absolute right for information as is recognised in the case of some court hearings, national security issues and Cabinet meeting minutes. While there should be a strong presumption in favour of releasing information about public institutions and others supported by government funding, these decisions should have regard to the public benefit versus public harm. In the case of school results, the evidence is that it brings no significant gains, but has significant negative effects on education and school communities while being subject to manipulation and rorting.

The right to information is a very important principle in a democracy. It is critical to keeping governments and government agencies accountable. However, there are some circumstances in which the provision of information can do greater harm than good. Reporting school results, and the inevitable league tables that follow, is one among many such circumstances.

Eminent overseas and Australian commentators on the use of performance indicators have acknowledged that publication of school results is not an unambiguous benefit to society and that there is a strong case for withholding such information. For example:

> As a reaction to unreasonable secrecy the belief in open access to information seems wholly healthy and undoubtedly has led to many benefits. Yet public disclosure of information cannot be upheld as an absolute principle. This is recognized by governments, for example, when they reserve the right to withhold information they deem to threaten national 'security'. Likewise, if publication of information in likely to harm individuals unfairly, or to mislead, then there is a case for refusing to publish. It is our contention that some published performance indicators which make statements about schools or other institutions fall into this category. Their capacity to reflect reality accurately may be extremely limited and their publication may cause inappropriate inferences to be drawn about institutions. [Myers & Goldstein 1997; see also Goldstein & Myers 1996]

> ...if the publication of certain information has the potential for harming individuals, or may be seriously misleading, then a justifiable case can be mounted for refusing its publication. It could be contended that much of what might be described as *educational performance indicators* based on measures of student achievement falls into this category. Its ability to reflect objective reality may be extremely limited, and its publication may therefore cause both misleading and incorrect inferences about schools and 'school effectiveness' to be drawn. In such circumstances, there is strong case for withholding publication. [Rowe 2000; see also 2004]

Decisions on the disclosure of information have long had regard to the public harm and benefit. This is widely acknowledged by governments, the courts and many organizations

with an interest in government information and data. The Prime Minister herself would be the first to argue that Cabinet documents and minutes should not be released to the public because it could inhibit or damage government processes. It is a pity that she will not recognize the damage releasing school results can do to the education of children.

Freedom of Information laws around the world contain provisions setting out categories of information that can be withheld from release [Banisar 2006; see also Independent Review Panel 2008]. There are a number of common exemptions found in nearly all laws. These include the protection of national security and international relations, personal privacy, commercial-in-confidence, law enforcement and public order, information received in confidence, and internal government discussions. Cabinet documents, for example, are kept secret for 30 years in Australia. These exemptions are based on the general assessment that public or private harm could be incurred which exceeds any benefits that may be obtained by releasing such information.

Courts also often apply restrictions on the release of information. In some types of case the courts issue restraining orders on publicity or, which may come to much the same thing, conduct some proceedings in private. This occurs when a trial in open court would render the proceedings nugatory. Examples are when the subject matter affects national security or is a trade secret. Another instance is proceedings concerning the welfare of a child, where a public hearing would undermine the object of the proceedings. For the same reasons the courts sometimes make orders restraining publicity concerning certain aspects of cases.

Now, publication of school results hardly falls within these typical exemptions to the principle of freedom of information. However, they serve to remind that this worthy principle is not absolute, as the rhetoric of the former Federal Education Minister implies.

Most national laws also include a “public interest test” that requires governments to balance the interest in withholding information against the public interest in disclosure. School results are the property of State and Territory governments which have explicit or implied statutory requirements for public interest tests for the release of information. Good public policy also demands that the costs and benefits of major new proposals be fully assessed.

The basic question at issue in deciding whether school results should be published is whether it will cause undue harm to some people and/or seriously mislead the public compared to the benefits. This elementary principle was not followed before the governmental agreement to publish school results. The Federal and State/Territory governments assume that transparency of government data on schools is absolute. There is no evidence that governments have properly assessed the public harm against the benefits. It was a decision taken on ideological and political grounds and not one based on evidence.

This imbalance in the harm and benefit of publishing school results demands a re-consideration by Australian governments. Its introduction was planned in secrecy without public consultation and debate. The Prime Minister wants schools to be open and transparent about their performance but she has been totally unwilling to subject her proposal to full public scrutiny.

6. Alternative approach

The major recommendation of this submission is that the My School site should be abandoned because of the harm it will do to education and because of the fundamental difficulty of comparing schools in a meaningful, reliable and useful way.

There is some case for publishing aggregated school results, without identifying individual schools. The results can be used to demonstrate whether the proportion of students not achieving satisfactory outcomes is general across the system or largely associated with particular schools and communities. Publication of these results can also inform the public as to whether average school results are improving over time, or not, and whether poor and good results are concentrated in a relatively few schools, and whether poor and good results are associated with aspects of student background. Publication of consolidated data of this kind could act as a performance marker for education systems and governments, because if differences associated with disadvantage are not being reduced, then it is clearly government policies that are failing.

Publication of the distribution of school averages across score ranges is sufficient to provide an accountability measure for systems and governments. The distribution of school results can be presented by indicating the number of schools whose average score falls within different ranges. This information could be summarised as a histogram for each strand and Year level for both literacy and numeracy.

While this submission argues that the My School approach is fundamentally flawed, and should be scrapped, there are some minor improvements which should be supported even by those who continue to support it.

At the very least, some key changes should be made. One change would be to dispense with the colour coding of school results because it contravenes the policies and protocols for reporting school results published by the national education ministers' council and because it is highly prejudicial to schools that are "red flagged".

As Federal Education Minister, Julia Gillard gave an absolute guarantee that *My School* would not "name and shame schools". Yet, this is precisely what it does. Schools with the lowest results are to be flagged 'red', meaning 'red for danger'. Many schools serving the most disadvantaged communities in Australia have been given red flags on My School. It is a signal to parents to keep away. This is public shaming of the worst possible, and most socially noxious, kind.

Giving schools a 'red flag' is designed to punish. Why else was the colour red chosen for these schools? By 'red flagging' schools, the Government is aiding and abetting an annual ritual hunt for the worst performing schools around Australia that happens in England and the United States. It unfairly condemns schools in the most difficult circumstances and makes their task harder. Punishing schools by publicly labelling them as 'failures' in this way is not the path to school improvement. It is likely to be counter-productive by undermining teaching and learning.

Another area where some improvement could be made to My School to reduce the scope for misleading parents and the public is to report the margin of error on each school results for each subject tested at each Year level. The current approach is inadequate as it does not

report numbers participating in tests. It does not meet requirements of MCEETYA policies and protocols.

However, this reporting would need to be accompanied by very strong caveats warning the public of the implications of the margins for error for comparing school results. There will be a strong temptation for parents to conclude that, say, a 30 point difference in results between two small schools means one is better than another, despite the fact that the two results are not statistically distinguishable.

7. Recommendations

1. The My School site should be abandoned because the data it provides is likely to lead to invalid comparisons of schools because of fundamental statistical issues concerning test results, as well as deficiencies in background information. This may lead to parents making choices which can undermine schools doing a good job in disadvantaged communities. In addition, the focus on literacy and numeracy is likely to lead to narrowing of the curriculum and teaching.
2. The costs of the proposed changes to data collection for My School should be considered because they are unlikely to make the comparisons on the My School site much more reliable.
3. Consideration should be given to collection and reporting of data which is consolidated in a way that avoids identifying schools, but which reports on how well systems and governments are achieving equity goals in relation to educational outcomes. Publication of a simple histogram of the number of schools in each state/territory whose average NAPLAN results for each domain and Year level tested fall within different score ranges should be considered. This does not require individual schools to be identified.
4. At the very least, two important changes should be make to the My School website:
    - The colour coding of individual school results should be abandoned because it is highly prejudicial; and,
    - The number of students participating in each test at each Year level in each school should be published so that the margin of error can be ascertained in each case.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA) 2010. Proposals for Enhancing the My School Website. 21 June. Available at: http://www.acara.edu.au/verve/_resources/ACARA_-_Update_on_changes_to_the_My_School_website_21.06.10.pdf

Banisar, D. 2006. *Freedom of Information Around the World: A Global Survey of Access to Government Information Laws*. Privacy International, London. Available at: http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-543400&als[theme]=FOI%20Home%20Page

Goldstein, H.& Myers, K. 1996. Freedom of Information: Towards a Code of Ethics for Performance Indicators. *Research Intelligence*, July, 12-16. Available at: http://www.cmm.bristol.ac.uk/team/HG_Personal/Full%20Publications%20-%20download/Table%20of%20publications.htm

Independent Review Panel 2008. *Enhancing Open and Accountable Government*. Discussion Paper. Review of the Freedom of Information Act 1992 (Queensland), January. Available at: http://www.foireview.qld.gov.au/

Levitt, S. 2007. More Evidence on the (Lack of) Impact of School Choice. *New York Times*, 4 October.

Lubienski, C. 2009. Do Quasi-Markets Foster Innovation in Education? A Comparative Perspective. Education Working Paper No. 25, Directorate for Education, Organisation for Economic Co-operation and Development, Paris, August. Available at: http://www.oecd.org/dataoecd/41/6/43415929.pdf

Myers, K. & Goldstein, H. 1997. Failing Schools in a Failing System. Institute of Education, University of London. Available at: http://www.cmm.bristol.ac.uk/team/HG_Personal/papers_for_downloading.htm

National Assessment Program for Literacy and Numeracy (NAPLAN) 2009. *National Report: Achievement in Reading, Writing, Language Conventions and Numeracy 2009*. Available at: http://www.naplan.edu.au/verve/_resources/NAPLAN_2009_National_Report.pdf

Rowe, K. 2000. Assessment, League Tables and School Effectiveness: Consider the Issues and 'Let's Get Real', *Journal of Educational Inquiry*, 1 (1): 73-97.

Rowe, K. 2004. Analysing and Reporting Performance Indicator Data: 'Caress' the Data and User Beware! Paper presented at the Public Sector Performance and Reporting Conference, Sydney, April. Available at: http://www.acer.edu.au/documents/Rowe-IR_Conf_2004_Paper.pdf