

# **Published School Results and League Tables are Misleading and Unreliable**

**Trevor Cobbold**

**SAVE OUR SCHOOLS**

**Research Paper**

**June 2010**

**<http://www.saveourschools.com.au>**

# Table of Contents

Summary .....	3
Differences in student composition influence school results .....	3
Other outside factors influence school results .....	3
School results are a selective measure of education .....	4
School results are subject to manipulation and rorting .....	4
School results are subject to statistical error .....	4
Little prospect that the reliability of My School can be improved .....	5
1. Introduction.....	6
2. Differences in school composition influence school results.....	6
2.1 Differences in socio-economic composition .....	6
2.2 Differences in ethnic composition .....	7
2.3 Other differences in student composition .....	8
2.4 Policy contradictions .....	9
3. Other outside factors influence school results .....	10
4. Published school results are a selective measure of education .....	11
5. School results are subject to manipulation and rorting.....	11
5.1 Reduced participation of low achieving students.....	12
5.2 Controlling school admissions to enrol higher achieving students .....	16
5.3 Cheating .....	18
5.5 Extending the school year.....	22
5.5 Manipulation of school results is a feature of high stakes testing .....	22
6. School results are subject to statistical error.....	23
6.1 Measurement and sampling error in test results .....	23
6.2 School reporting should be statistically valid and reliable .....	26
7. Conclusions.....	26
References.....	30

## 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 reporting; 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 be 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 cheating**

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 sorting; 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 9<sup>th</sup> grade for 2 years or more and then suddenly reappeared in the 12<sup>th</sup> grade. This practice had two benefits for schools. First, by skipping 10<sup>th</sup> grade, students did not take the state test in the year that it counted for school accountability ratings. Second, by showing up in 12<sup>th</sup> grade, they contributed to a more favourable rating where school progression is examined as the proportion of 9<sup>th</sup> graders who appear in 12<sup>th</sup> 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 Illinois 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 their 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 25<sup>th</sup> to the 75<sup>th</sup> 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 distort 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 distortion continue to be a feature of systems that publish school results and league tables.

Manipulation and distortion 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 distinguishable.

## 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](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](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](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](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](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](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](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](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](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](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](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](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](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=ViewSinglePubl ic&LinkID=211&ModuleID=38&NEWSPID=1](http://www.performanceincentives.org/news_events/detail.asp?pageaction=ViewSinglePubl ic&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](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](http://www.statscom.org.uk/C_144.aspx)
- Suliman, R. & McInerney, 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](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](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/](http://www.rand.org/pubs/rgs_dissertations/RGSD256/)