

Reporting School Results and Student Achievement

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

Part 1

SAVE OUR SCHOOLS

June 2010

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

Table of Contents

Key Points.....	3
Summary.....	4
1. Introduction.....	6
2. School report card studies.....	7
3. High stakes accountability studies.....	7
3.1 Pre-NCLB studies.....	8
3.2 NCLB studies.....	11
3.3 Impact on the distribution of student achievement.....	12
3.4 Cross-country studies.....	14
3.5 Australian Government claims.....	15
4. Choice and competition studies.....	16
5. Conclusions.....	19
References.....	21

Key Points

1. A review of the major research studies on reporting school results, school accountability measures and policies to promote greater choice and competition between schools shows that the Government's claim that publishing school results will improve student achievement is unsubstantiated. At best, the evidence is mixed. This is not a robust foundation for education policy.
2. Studies of the impact of reporting school results show show no significant effect on student achievement.
3. The evidence from the major studies of "high stakes" accountability measures (which include reporting school results, rewards and sanctions for schools, exams for students and grade promotion standards) is mixed. Some find positive effects on student achievement, others find no effect or negative results.
4. The most frequently cited studies showing a positive effect from "high stakes" accountability over-state the effect because of technical flaws in the measurement of the effect.
5. Much of the positive effect of school accountability measures shown in key studies is due to student-based requirements such as meeting grade promotion standards, passing end-of-course exams and graduation exams. The effect of these requirements on student achievement is larger than the effect of rewards and sanctions relating to school performance.
6. There is substantial evidence that the positive results in some studies may be due, at least in part, to schools manipulating their results in various ways such as re-classifying students as special education students so they can be exempted from high stakes tests, suspending low achieving students during the testing cycle and outright cheating.
7. The major academic reviews of "high stakes" accountability studies conclude that the evidence is mixed and provides little scientific foundation for these policies. This has been acknowledged by the chief executive of the Australian Curriculum, Assessment and Reporting Authority, Peter Hill, thus contradicting his Minister's claims.
8. The weight of evidence from the best designed and most comprehensive of studies on increasing choice and competition between schools is that it does not improve student achievement once student background characteristics are taken into account.
9. Former advocates of "high stakes" school accountability measures and greater choice and competition between schools such as Chester Finn and Diane Ravitch have been forced to admit that these policies have not worked.

Summary

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

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

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

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

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

There are several reasons to believe that the positive results shown in some studies are overstated.

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

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

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

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

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

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

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

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

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

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

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

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

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2. School report card studies

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

School report cards and other public reporting strategies appear not to have significant or consistently positive relationships with the NAEP test scores.

Public reporting is necessary for the implementation of these other policies [rewards and sanctions], but on its own it had no discernible effect on student achievement. [Bishop et.al. 2001: 293, 310]

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

The analysis, however, indicates that just reporting results has minimal impact on student performance....

States that simply provide information through report cards without attaching consequences to performance do not get significantly larger impacts than those with no accountability. [Hanushek & Raymond 2005: 298, 321]

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

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

3. High stakes accountability studies

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

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

3.1 Pre-NCLB studies

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.2 NCLB studies

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

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

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

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

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

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

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

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

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

3.3 Impact on the distribution of student achievement

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

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

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

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

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

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

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

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

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

3.4 Cross-country studies

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

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

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

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

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

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

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

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

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

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

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

3.5 Australian Government claims

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

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

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

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

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

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

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

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

4. Choice and competition studies

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

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

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

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

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

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

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

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

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

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

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

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

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

The study concluded:

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

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

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

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

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

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

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

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

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

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

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

Choice and accountability, I believed, would offer a chance to poor children to escape failing schools. Testing and accountability, I thought, would cast sunshine on low-performing schools and lead to improvement...

Today there is empirical evidence, and it shows clearly that choice, competition and accountability as education reform levers are not working. [Ravitch 2010b]

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

5. Conclusions

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

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

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

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

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

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

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

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

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

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

References

- Allen, R. & Vignoles, A. 2009. Can School Competition Improve Standards? The Case of Faith Schools in England? Paper presented to the Annual Conference of the Royal Economic Society, April. Available at: <http://repec.ioe.ac.uk/REPEc/pdf/qsswp0904.pdf>
- Amrein-Beardsley, A. & Berliner, D. 2003. Re-analysis of NEAP Math and Reading Scores in States With and Without High-stakes Tests: Response to Rosenshine. *Education Policy Analysis Archives*, 11 (25).
- Ballou, D. & Springer, M. G. 2009. Achievement Trade-offs and No Child Left Behind. Paper presented to the *NCLB: Emerging Findings Research Conference* at the Urban Institute, Washington, DC, 12 August 2009. Available at: <http://www.caldercenter.org/nclbconference.cfm>
- Bifulco, R. & Bulkley, K. 2008. Charter Schools. In Ladd, H.F. & Fiske, E.B. (eds.). *Handbook of Research on Education Finance and Policy*. London: Routledge.
- Bishop, J.; Mane, F.; Bishop, M. & Moriarity, J. 2001. The Role of End-of-Course Exams and Minimal Competency Exams in Standards-Based Reforms. In Diane Ravitch (ed.), *Brookings Papers in Educational Policy*, Brookings Institution, Washington DC: 267-344.
- Braun, H. 2004. Reconsidering the Impact of High-Stakes Testing. *Education Policy Analysis Archives*, 12 (1), January.
- Bruniges, M. 2009. Opening Address. Australasian Curriculum, Assessment and Certification Authorities Conference. University of Western Australia, Perth, 13 July.
- Carnoy, M. & Loeb, S. 2002. Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24 (4): 305–331.
- Center for Research on Education Outcomes (CREDO) 2009. Multiple Choice: Charter School Performance in 16 States. Stanford University, Stanford, CA, June. Available at: <http://credo.stanford.edu/>
- Cobbold, T. 2007. The Great School Fraud: Howard Government School Education Policy 1996-2000. Available at: <http://www.saveourschools.com>.
- Commonwealth of Australia 2008. Boosting Australia's Productive Capacity: Infrastructure Skills. *Statement No. 4, Budget Paper No. 1: Budget Strategy and Outlook 2008-09*. Available at: <http://www.budget.gov.au/2008-09/content/bp1/html/index.htm>
- Cullen, J & Jacob, B. 2009. Is Gaining Access to Selective Elementary Schools Gaining Ground? Evidence from Randomized Lotteries. In Jonathan Gruber (ed.), *An Economics Perspective on the Problems of Disadvantaged Youth*. National Bureau of Economic Research, University of Chicago Press, Chicago, 43-84.
- Cullen, J.; Jacob, B. & Levitt, S. 2005. The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools. *Journal of Public Economics*, 89 (5-6): 729-60.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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