

Professional Voice

Volume 10 Issue 1 | Winter 2014 | RRP \$10

ISSN: 1445-4165

TESTING times



Editor

John Graham

Editorial Committee

John Graham, Meredith Peace and Nic Barnard

Designer

Peter Lambropoulos

Published by

Australian Education Union Victorian Branch

Printed by

Printgraphics

Printed on recycled paper



Cover: 310g ReArt Gloss

Text: 105g ReArt Matt

Annual subscriptions

\$25 within Australia (free to financial AEU members). Please contact the AEU Publications Unit on 03 9417 2822 for further information. A subscription form is available in the back of this issue or downloadable from www.aeuvic.asn.au/publications.

Back issues

Free to members. \$10 to non-members within Australia.

Advertising is accepted according to AEU guidelines. Please contact Lyn Baird for further details. Phone: 03 9418 4879.
Email: lyn.baird@aeuvic.asn.au.

Editorial correspondence should be directed to Nic Barnard, Professional Voice, Australian Education Union Vic Branch, PO Box 363, Abbotsford VIC 3067.
Phone: 03 9418 4841.
Email: nic.barnard@aeuvic.asn.au.

Copyright is held by respective authors. The opinions of the authors are their own and are not necessarily the policy of the Australian Education Union Victorian Branch. For authorisation to reproduce any article in part or whole, please contact the editor.

The AEU Victorian Branch can be contacted at:
PO Box 363, Abbotsford 3067
Tel: 03 9417 2822 Fax: 1300 568 078
Web: www.aeuvic.asn.au

ISSN:1445-4165

Testing times



Professional Voice

Volume 10 Issue 1

Contents

- 5 Editorial: Testing times
John Graham
- 10 Flaws in the lens
Alan Reid
- 16 Linking student test scores to teacher performance
Margaret Wu
- 24 A cautionary tale
David Hursh
- 30 PISA 2012: Australia's performance
Sue Thomson
- 41 The importance of class size
David Zyngier
- 47 Pasi Sahlberg on lessons from Finland
Interview by John Graham

Editorial: Testing times

John Graham

Arguably the most prominent education policy development in the 21st century has been the rise and impact of international testing. Beginning in the year 2000, the OECD's PISA program in particular has set up a sort of education Olympics where countries compete against each other to rise to the top of published league tables. The position each country has on the league table, and whether it's rising or falling, has become (to continue the sporting metaphor) a political football that governments and opposition parties both use to score political points and justify their policies. While there is no proven link between PISA success and economic success, the assumption that the two are intimately connected in the race for prosperity in a globalised world permeates the political discourse about education. It also invades the dreams of economists.

The conservative American economist Eric Hanushek responded to the release of the 2012 PISA results by contending that if the US could attain the same level as Canada it "would translate into 20% higher paychecks for the average worker over the entire 21st century." An earlier Australian example of the same argument was put forward by Ben Jensen from the Grattan Institute in 2010 using the PISA 2006 results. He argued that Australia could sit at the top of the PISA table with Finland, Hong Kong and Canada if it increased teacher effectiveness by 10% and student learning by 5% pa. This in turn would improve the productivity of Australian workers and "increase long-run economic growth by \$90 billion by 2050, making Australians 12% richer by the turn of the century."

When the 2012 PISA results were released in December 2013 and it was found that Finland's pre-eminence had been eclipsed by a suite of Asian countries, its media lamented that "the golden days are over" and concern was raised by its national broadcaster that its education system may suffer the same decline as Nokia. The Finnish Minister of Education and Science talked about taking "strong action" by bringing in experts in research and education to address the downturn. In the UK, the government came under pressure

John Graham is a research officer at the Australian Education Union (Vic). He has carried out research in a wide range of areas related to education and training. He has had particular responsibility for the range of issues impacting on teachers and teaching as a profession, teacher education, developments in public education in general and in schooling in particular, the organisation and funding of education and curriculum change. He has written extensively about all of these matters. This includes many articles, submissions to state and federal inquiries and reviews, development and analysis of surveys and various papers to seminars and conferences.

because the country's performance was "stagnating" and countries from which they had borrowed their reforms (such as "free schools" from Sweden) were now rapidly dropping down the league tables. The UK Education Minister, Michael Gove, said their "poor" results meant there was an urgent need to speed up the implementation of more rigorous exams and more academies and free schools (privately run schools within the public system) which were, coincidentally of course, their existing policies. And, because Shanghai-China topped the rankings, Chinese maths teachers would be invited to come to Britain to give "master classes" to British teachers in how to teach the subject.

In Australia the gloom and doom reaction to the PISA 2012 results was all about the country's "further" decline in performance. To Christopher Pyne, the Federal Minister for Education, however, (who unlike Michael Gove had just come to power) they looked more like a gift from the gods. When in Opposition he had used previous results to argue that they showed Australia had "wasted" money on education (especially through class size reductions) because the country's performance was falling rather than rising in the international league tables. The fact that for eleven years he was a member of the former Howard Government was never mentioned. Now in government again, he saw the 2012 results as further evidence that Australia needed to take a new direction in education. The results were "the worst for Australia since testing began" and showed Labor's Education Revolution was "a spectacular failure". He claimed that the results proved that "more funding does not equate to better outcomes" and that the country did not have a problem with equity. He also claimed that the results showed the rightness of his government's policy directions – the need for a curriculum review, greater school and principal autonomy, better classroom pedagogy and better quality teachers – "because these are the things we know will lift results".

In other words, the bad news about Australia's PISA performance looked like good news for the new minister. There was no need for him to take responsibility for the results. It provided ammunition against his political opponents in the previous government, it provided justification for not implementing policies his government did not support (like Gonski), it created a supposed need for the new policies the Abbott Government was wanting to implement, and it could be used in 2014 to make cuts to education and claim that they would have no impact on student achievement as education funding and success were apparently unrelated.

Almost all of Minister Pyne's interpretation of the PISA results is misleading. It entails a wilful misreading of the data and other OECD research findings. The evidence that Australia has an equity problem is clear from any comparative analysis of published PISA data (see Sue Thomson's article in this edition of *Professional Voice*). Particularly relevant to Gonski,

Australia is shown as having one of the most inequitable systems of resource allocation of any country in the OECD. The OECD does not say that funding has no impact on achievement; it indicates that the relationship is complex because of the various factors that influence performance in each country.

Schools that have greater autonomy over curriculum and assessment, and which have more teacher participation in school management and greater teacher-principal collaboration, perform better. There is no evidence that more autonomy for schools over resources and staffing (the Coalition model) has any impact. There is also no reference whatsoever in the OECD data to a link between the imposition of a culturally conservative curriculum and improved student outcomes. As for teacher quality and pedagogy, there is no credible documented evidence in Australia that teacher quality is declining or that the pedagogy used in public school classrooms is deficient.

The Victorian Coalition Government's response to the 2012 PISA results was more muted as it had already hitched its "New Directions" policies to the 2009 results. The *New Directions* paper (2012) leveraged a package of radical GERM policies using the supposed mediocre performance of the state's 15-year-old students in PISA reading tests from 2000 to 2009. The performance was described as "good but not great" and Victoria needed radical school reform to climb into what the paper called "the global top tier". It even set out a precise formula to achieve that aim. The state's 15-year-olds needed to progress on average an extra six months in their learning by the time they entered Year 10. They would only do this if a range of policy initiatives were implemented, including "exiting" the lowest performing 5% of teachers, improving teacher productivity by 2%, bringing in managers from industry to be principals, introducing performance pay, and watering down teacher registration requirements.

While the paper claimed that its package was a golden staircase to the top of the PISA rankings, it contained no reliable evidence base to back this up, and no analysis showing the "global top tier" systems had followed a similar path. The paper's formula was little more than an ideological wish-list masquerading as a summary of world's best practice. PISA provided the data to give this wish-list a skerrick of policy respectability.

Disquiet over the use (or rather misuse) of PISA data has led to a worldwide group of over 100 education academics and other key education stakeholders writing to the OECD, calling for the overhaul of the PISA testing program because of the negative consequences of league table rankings. The letter outlined the many world-wide negative effects of the testing program, including an escalation in standardised testing; short-term fixes to enable a country to quickly climb the rankings; emphasis on a narrow range of measurable

aspects of education; and the “dangerous narrowing of our collective imagination regarding what education is and ought to be about”. It criticised the OECD, with its narrow focus on standardised testing, for assuming the power to shape education policy around the world and “risk turning learning into drudgery and killing the joy of learning”.

The articles and the interview in this edition of *Professional Voice* are all related in some way to the impact of testing on students, teachers and schools. In our extended interview with Pasi Sahlberg about education in Finland and its implications for education systems in Australia, there is much about the use and misuse of PISA data. Finland came to prominence in education circles around the world because it topped the PISA performance tables. According to Sahlberg, this created more interest in other countries than in Finland itself. He is critical of those countries like Australia which have narrowed their educational vision to focus on becoming top of the PISA global league tables. This has the potential to sell students short by concentrating on a limited area of academic achievement at the expense of broader learning and personal development goals.

On the other hand, Sahlberg believes PISA has played a key role in putting equity on the global education agenda. The most successful PISA performers are those with high equity as well as high learning outcomes. This combination is at the heart of the Finnish school system and is central to its PISA performance. He is also clear that his country achieves these results through its strong public education system.

In his analysis of the Abbott Government’s education policies, Alan Reid agrees with Pasi Sahlberg that PISA should not be used “as the benchmark for our national educational aspirations”. He is particularly concerned that the Federal Government is relying on PISA — a set of survey results at one stage of schooling every three years in only three areas of the curriculum — as evidence that Australian education standards are falling. This approach means that problems with the country’s education system are misdiagnosed and the “reforms” to fix them may damage rather than improve the work of schools. Reid believes that the most egregious claim by the Federal Minister for Education is that Australia does not have an equity problem. He sees this position as a device to justify not shifting resources from the most to the least advantaged and not properly implementing the Gonski reforms.

Sue Thomson, who is one of the authors of the ACER report on Australia’s PISA performance, provides a summary of the 2012 results and the issues that arise from them. The results indicate that the achievement of Australian students in mathematical literacy and reading literacy, while still above the OECD average and relatively high in world terms, declined in 2012, lowering the proportion of 15-year-olds at the higher performance levels and increasing the proportion at the lower levels compared to previous test years. She also

sets out the wide range of data, which substantiates the case that Australia has an equity problem requiring urgent attention.

Margaret Wu and David Hursh separately write about the misuse of standardised testing by governments and education authorities in Australia and the United States. Wu, who is an expert in educational measurement, sets out an evidence-based case against the use of NAPLAN results to measure individual teacher performance. It is neither valid nor reliable to evaluate teachers using standardised test data, and attempts to do so end up damaging the teaching profession and education more generally. Hursh describes the politics of high-stakes testing in New York and the way it has been used to undermine the public education system there. His compelling picture of politicians bending the truth to further their agenda of privatisation sounds a lot like the first nine months of the Abbott Government.

Our final article is about one of the most contested areas of education research – the effect of class size. Christopher Pyne has said on many occasions (using PISA data) that small class sizes are a waste of money and that there is no research showing that they improve student outcomes. David Zyngier recently completed the first thorough review of class size research by an Australian academic for some time. He reviewed 112 research studies and found that the evidence strongly indicated that reduced class sizes made a measurable difference to student achievement, particularly for children from disadvantaged backgrounds.

References

- Hanushek, Eric A., (2014, January), *US Results on PISA 2012*, <http://hanushek.stanford.edu/sites/default/files/publications/PISA%20overview.jan2014.pdf>
- Jensen Ben, (2010), *Investing in our teachers: investing in our economy*, Grattan Institute, <http://grattan.edu.au/publications/reports/post/investing-in-our-teachers-investing-in-our-economy/>
- Meyer Heinz-Dieter and Zahedi Katie, and signatories, (2014, May), *An Open Letter: To Andreas Schleicher, OECD, Paris*, 5th Global Policy Journal <http://www.globalpolicyjournal.com/blog/05/05/2014/open-letter-andreas-schleicher-oecd-paris>
- OECD (2013), *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices Volume IV* <http://www.oecd.org/pisa/keyfindings/Vol4Ch1.pdf>
- Pyne Christopher, (2013, 3 December), *PISA results show more work to be done*, Media Release, <http://ministers.education.gov.au/pyne/pisa-results-show-more-work-be-done>

Flaws in the lens

The Abbott Government's assumptions about education

Alan Reid

The **ideological lens** through which the Coalition sees its education policy frames everything around the individual rather than society. But inside this broad ideological framing, the Government needs to establish pragmatic reasons to justify radical change to the public education system. To achieve this, its policy platform appears to be based on a set of oft-repeated assumptions.

Thus, to understand government education policy, it is necessary to disentangle the assumptions from the policies they justify. Two of the most prominent assumptions are that Australian education standards are generally declining and that equity is not a problem in Australian education. I will argue that since these assumptions do not stand up to critical analysis, the policy which is based upon them is highly problematic.

Assumption 1: That Australian education standards are generally declining.

How many times have you seen a media opinion piece, a government report, a think-tank education report, another politician or a letter to the editor start with the claim that standards in Australian education are declining compared to other developed countries? Here's an example drawn from an Alan Jones interview with Education Minister Christopher Pyne:

Pyne: ... In the last 10 years our results have been declining. So we've spent 40% more on school education in the same time that our results had declined, not just in relative terms against other nations, but in real terms. ... That is a complete failure of schooling when you are spending more money and you're going backwards in real terms and the reason is because of what we're teaching our children and the how we're teaching our children.

Alan Reid is Professor Emeritus of Education at the University of South Australia. He is involved in a range of national and state professional organisations. His research interests include educational policy, curriculum change, social justice and education, citizenship education and the history and politics of public education. He has published widely in these areas and his contribution to education has been recognised by a number of national awards. He is also involved in policy development at the state and national levels.

This sort of assertion is presented as self-evident, not a matter of opinion requiring evidence. Once accepted, policy consequences follow. Thus, in this example, Minister Pyne is now in a position to imply that we have been spending too much money on schools, and to diagnose that the cause of the decline is what and how we teach our students.

If challenged to produce the evidence that shows that educational standards are falling, invariably it is Australia's performance on the Programme for International Student Assessment (PISA) tests which is proffered. It is therefore pretty important to know if the test results really are a good measure of education quality.

PISA is an OECD-administered test of the performance of students aged 15 in maths, science and reading. It has been conducted every three years since 2000, with the 2012 results released in December 2013. PISA purports to test "real-life" skills and knowledge. About 65 countries participate in the tests which are two hours in length and hand-written, and involve a stratified sample of students in each country. In 2009 Australia was ranked equal 10th in reading, equal 17th in maths and equal 8th in science.

Press commentators and politicians, when PISA results are published, blithely ignore the warnings on the PISA site not to use the test scores to make superficial judgments comparing the quality of education in various countries. The "winners" are eulogised while those countries which have slipped a few rungs on the league table are excoriated. What is said confirms the public perception that PISA is a scientific and objective measure of education quality. This has a number of consequences which I will call the PISA effect.

In recent years when Australia has slipped a few rungs, there has been a storm of media attacks on educators and policy makers, and a flurry of favourite policy positions proposed to address the decline, such as performance pay for teachers, greater school autonomy, revamped teacher education programs, and voucher systems to enable school choice. It rarely seems to trouble the commentators that there is no evidence to support a relationship between the PISA data and the proposed solutions.

The situation has also created a sense of educational crisis with schools and teachers bearing the brunt of criticism, with a flow-on negative impact on morale. It has spawned a new education industry involving visits to the top five countries to discover why they are more successful than Australia. This "research" consists of cherry-picking some of the structures, practices and processes of the league table leaders, guessing which variables have contributed to their success, aggregating the variables, and then urging that these practices be adopted in Australia. A classic of this genre is the Grattan Institute's 2012 report *Catching up: Learning from the best school systems in Asia* (see Reid, 2013).

These various PISA effects provide the justifications needed for policy makers to push their favoured policies. However, there are a number of problems with blindly accepting the PISA outcomes. The first is that many technical questions are being asked about the efficacy of the PISA tests and their associated processes. Some of these relate to:

- Cultural factors such as the difficulties associated with making an international test which is converted to many languages culturally neutral, and the propensity of some systems to prepare students for the test (such as the extensive use of after-hours tutors)
- Vast differences in contexts between countries, and the fact that other factors are ignored. In Singapore, for example, there is a concern that although students are successful in tests, their creativity is stifled by a narrow and strait-jacketed curriculum. Uncritically importing policies and practices from other countries is fraught with danger, particularly when the only reference point is an international test
- Selective sampling of some systems such as Shanghai where the majority of students from disadvantaged rural backgrounds are excluded; indeed China has not allowed the OECD to release the results of a number of other provinces where students sat the test
- Increasing doubts raised by education researchers about the validity and reliability of the tests, for example because the items compared are not all the same across all countries
- The difficulties associated with ascribing causality to cross-sectional data, which can only be resolved by exploring the relationship between different variables using longitudinal models.

However, it is from the perspective of the public good that the reliance on PISA is most troubling. Surely results at one stage of schooling every three years in only three areas of the curriculum are too narrow a base upon which to make claims about the quality of Australian education. The fact is that although reading, maths and science are important, they tell us nothing about outcomes in such crucial areas as the arts, history, geography, health and PE, English literature, civics and citizenship, and information and communication technology, to name just a few areas of the formal curriculum. More than this, such a narrow view of education means that we get no sense of how students are faring in such critical domains as problem-solving, inquiry, creativity, intercultural understanding, and interpersonal relationships. At best the international test results present a very limited picture of student progress; and certainly do not include the many characteristics that add such strength to Australian public schools.

I am not saying that PISA is a shoddy test; nor am I saying that such international tests cannot tell us anything. Rather I am making the point that policy, media commentary and

research which is premised on PISA results should at least acknowledge these difficulties and limitations and be much more tentative about using PISA as the sole arbiter of what constitutes quality in education. The use of PISA as the benchmark for our aspirations is fraught with danger. We need far more extensive and sophisticated measures which take account of the public purposes of education.

This might seem obvious, and yet the view of PISA as an objective measure of education quality has become so dominant in the public arena that it is no longer questioned. PISA outcomes appear to be the major rationale for the government's education policy agenda.

In short, the assumption that standards in Australian education are falling generally does not stand up to critical scrutiny. I want to stress that I am not arguing that all is well in Australian education. Indeed, the equity gap in Australia is a major concern, as I will show in the next section. But it does suggest that we should be prouder and more supportive of the work done every day in Australian schools. My fear is that imposing a range of "reforms" to fix a misdiagnosed problem on the basis of a narrow data set is more likely to damage than enhance the great work that is already occurring in our schools, and at the same time serve to widen rather than diminish the equity gap in Australian education.

Assumption 2: That equity is not a problem in Australian education

In November last year, Minister Pyne, told the *Lateline* program categorically:

I don't believe there is an equity problem in Australia.

He went on to argue that the Gonski reforms had nothing to do with equity. This of course explains a lot about the Federal Government's educational priorities, particularly its backsliding and broken promises on funding. We have reached the stage where our federal education minister is willing to publicly deny 30 years of evidence about inequalities in the Australian education system. It is a false assumption which needs to be exposed.

There are any numbers of indicators which show that equity is a very real issue, and it is incredibly sad that we need to spend time on demonstrating its existence, rather than working on approaches to address it. But here we go again.

Let's start by using the PISA data which many, including Minister Pyne, cite to support the claim that standards are declining. The recently released 2012 PISA results show large gaps between the outcomes of students from educationally advantaged and disadvantaged backgrounds. That is, at age 15, Indigenous students are three-and-a half or more years

behind high SES students, and remote students are three years behind. Rural students and students from low socio-economic backgrounds are heavily represented in the lower levels of the test results.

But of course there are many other indicators of the equity problem in Australian education. For example, analyses of the annual NAPLAN scores regularly demonstrate the large gap between the results of students from educationally advantaged and disadvantaged backgrounds. The results demonstrate distinct patterns of disadvantage based on race, gender, socio-economic background and location. A UNESCO report published in January looked at data between 1994 and 2011 and concluded that 16% of the poorest year 8 students in Australia are below the minimum benchmark in maths, compared with just 4% of students from the most advantaged backgrounds. Many studies have shown that at the senior end of high school, students from educationally disadvantaged backgrounds are poorly represented in the numbers reaching year 12, and in achieving a senior school certificate, and are heavily under-represented in high ATAR scores in high-status academic subjects (eg Teese, 2005). In addition, Australian schools in high socio-economic areas are better resourced generally than those in low socio-economic areas.

Public schools educate the vast majority of students who are educationally disadvantaged and/or have special needs. Thus public schools educate 80% of students in the lowest SES quartile (and the percentage is growing); 85% of Indigenous students; 83% of students in remote/very remote areas; nearly 80% of students with a funded disability; and the vast majority of students with English language difficulties. It follows that most of the heavy lifting involved in supporting these students is borne by public schools. It would seem axiomatic then that if a government wanted to narrow the equity gap, it would provide additional resources and support to the schools which are bearing the burden of the serious educational challenges involved in supporting these students.

But no, it seems that Minister Pyne finds it much easier to say that Australian education does not have an equity problem, rather it has a "student outcomes problem" (*Lateline* July 16, 2012). This tautology perhaps points to the thinking about equity which "informs" the minister. Recently, his close adviser Dr Kevin Donnelly wrote a very revealing piece in the *Weekend Australian* (December 28–29, 2013) arguing that innate cognitive ability has the greatest influence on student outcomes. At first blush it seemed that he was explaining that Indigenous and low-SES students had inferior educational outcomes because they are not as intelligent as those from more affluent backgrounds. When he then quoted Charles Murray, the controversial American academic, it became clear that this was precisely what he was arguing.

Murray and Hernstein published the book *The Bell Curve* 20 years ago. Using IQ tests, they purported to show that certain groups in the US population such as Blacks were overrepresented among the less intelligent. They claimed that since intelligence is inherited rather than also being (perhaps primarily) the result of social and cultural factors, efforts to improve the educational and economic opportunities of poor people, especially black poor people, are doomed to fail. Since that time, much of the research upon which this thesis was posited has been discredited.

It is very sad, that after four decades of thinking, research and practice related to ways to reduce the equity gap in Australian education – a complex task where small gains have been made but much is yet to be achieved – we seem to have turned the clock back to the point that an old fashioned and discriminatory idea based on flawed science is seriously floated in the public arena by a key government education advisor.

Recognising the equity gap is not to say that all students should have the same learning outcomes, or that socioeconomic factors are the sole determinant of educational outcomes, as Donnelly has asserted that the “education establishment” argues. Rather it is to recognise that all students do not start with the same life chances or have access to the same resources. If we are serious about education as a public good, then we must seek to ensure that educational disadvantage is overcome in order to give *all* young people the opportunity to fulfil their individual potential.

To claim that there is no equity problem in Australia relieves the government of the burden of having to shape policy which shifts resources from the most to the least advantaged. This of course is the reason for Minister Pyne’s many policy positions on the Gonski formula. If there is no equity problem, then he can justify a hands-off approach to the distribution of funds on a needs basis, or not require that states and territories co-contribute by topping up federal money.

I have explored two dominant assumptions that underpin the government’s education policy. They have been repeated so often that they have almost become established fact. These and other similarly incorrect assumptions are the grounds upon which the government justifies specific policy initiatives. If, as I have argued, these assumptions do not stand up to critical scrutiny, then any policies which have apparently been designed to address them, become highly suspect.

Standardised testing has proliferated over the past ten years. More alarmingly, the high-stakes use of student assessment data has also increased. The publication of teacher rankings in the *New York Post* (February 2012) is an example of such inappropriate use of student data. This paper attempts to explain reliability and validity issues in using student test gain scores as teacher performance measures. In particular, for individual teachers, test mean scores for a class have large variability, making student outcome measures unstable for assessing teacher performance. This is the reliability issue. There are also many factors other than schooling affecting student performance so student outcomes may not reflect teachers' efforts. This is the validity issue.

Standardised testing has proliferated since the No Child Left Behind Act was introduced in 2001 in the United States. Further, assessment results from standardised tests have become high-stakes through, for example, the linking of test results to teacher performance measures.

Despite repeated warnings by academics (eg Darling-Hammond *et al*, 2011) about the inappropriate use of student test scores for teacher accountability, such uses still abound in government education departments and the media. The publication of the names of teachers and their "performance scores" by the *New York Post* in February 2012 (bit.ly/1oCVgzp) was so disturbing that discussions on the use of student assessment data for evaluating teachers ought to come to the fore.

One possible reason for the proliferation of the use of student assessments to measure teacher performance could be the accessibility of complex statistical analyses. With readily available software programs, statistical methodologies such as multilevel modelling, variance decomposition and, in particular, value-added models can tease out the contributions of

Margaret Wu has a background in educational measurement and statistics. She has worked as a statistician and a psychometrician for many years. Margaret has been closely involved in large-scale surveys including PISA, TIMSS and Australian national assessment programs. She has co-authored two software programs that have been used nationally and internationally for the analysis of assessment data. Margaret has written numerous academic papers and book chapters in the field of assessment.

various factors to student achievement. While sophisticated statistical models are great in helping us understand data, what has often been ignored is a check on the reliability and validity of the results. In particular, when the data collected lack accuracy, no statistical modelling can provide trustworthy results.

This paper explains the limitations of the use of student assessment data in measuring teacher performance. It does not dispute that teachers make a difference to student achievement. But many factors influence student achievement, making it extremely difficult to isolate individual teacher effects. While value-added models could provide a ballpark figure of teacher effect as a cohort, the estimation of individual teacher effect is fraught with problems. This is not a deficiency in value-added modelling, but in the data we put into the models. The main aim of this paper is to demonstrate that using student achievement data for teacher performance measures is far off the mark. The discussion below is organised around issues of reliability and validity of data and results. In the context of using student achievement to evaluate teacher performance, reliability refers to the accuracy of student achievement measures, and validity refers to the appropriateness of inferring teacher performance from student achievement.

Reliability issues

Measuring group effect versus individual effect:

Many education research studies use value-added models to estimate the effects of factors contributing to students' academic success, including teacher effect. With a large amount of data, we can estimate a ballpark figure for how teachers make a difference to student achievement. Teacher effect is found to be a little under one year of growth, where "one year of growth" refers to the average growth of students in one calendar year.

For example, Hattie (2008) showed that teacher effect is about 0.3 in effect size, where 0.5 is roughly one year of growth. That is, a very effective teacher can bring students up one-and-a-half years in one calendar year, while an ineffective teacher brings students up only half-a-year. These figures are plausible. Any larger teacher effect would not be credible. It would be difficult to find teachers who can make a whole class of students finish six years of primary schooling in three or fewer years, except perhaps for a handful of gifted students. To claim that a teacher can make a class of students grow two or more years in one calendar year is simply not likely. It's true that the press likes to report success stories, but these are mostly exceptions.

The fact that we can get a handle on an overall teacher effect leads many people to think that we can measure an *individual* teacher's effect accurately. This is a misperception.

We can often measure a group effect without being able to measure individuals accurately. For example, we want to know the average age of a group of people, but it may be impolite to directly ask their ages. So in a questionnaire we group ages in 20-year intervals instead of asking a person's actual age. We will not know each person's age accurately, but if the sample size is large enough we can obtain a fairly accurate average age. The same applies to measuring teacher effectiveness.

Since the range of teacher effect is about one year of growth, to separate effective from ineffective teachers our measures need to be accurate to fractions of a month's growth. Achievement tests just don't provide that kind of accuracy. Wu (2010) showed that based on measurement error alone (uncertainty due to the shortness of a test instrument), the 95% confidence interval for a class mean test score of 30 students is about 0.3 in effect size. Adding sampling error (variation due to different cohorts of student from one year to another), it is conservative to say that the margin of error surrounding a performance measure based on a class mean gain score is about as large as the size of the total teacher effect. Figure 1 shows this pictorially. Even if there are multiple years of student assessment data, the reduction in the margin of error is still nowhere near acceptable for deciding on teacher performance.

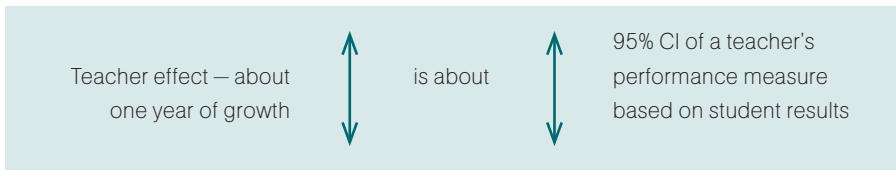


Figure 1: Comparison between the size of teacher effect and the accuracy of teacher performance measure

Error rates in measuring teacher performance using student test score gains:

A report from the US National Centre for Education Evaluation (Schochet & Chiang, 2010) concludes that the error rate in labelling an average teacher as high or low-performing is one in three, when one year of student test gain scores are used. When three years of data are used, the error rate is one in four. Note that one cannot do any worse than an error rate of one in two, since that is the outcome of tossing a coin. Essentially, the accuracy of labelling teachers as effective or ineffective based on students' test gain scores is marginally better than tossing a coin. The reason for this is that teacher effect is small in comparison to the variability in student achievement. Teacher effect accounts for about 10% of the variations in student achievement (eg Leigh, 2010).

In the report *Evaluating Teacher Evaluation* (2011), Darling-Hammond and colleagues noted that many teachers observed that their class results varied greatly from one year to

another when they had not changed their teaching methods. McCaffrey, Koretz, Lockwood, Louis & Hamilton (2004) also warned about the lack of precision in using value-added models to rank teachers based on student test scores.

The *New York Post* article gave teachers percentile ranks between 0 (low) and 99 (high) based on student performance (where a percentile rank of 65 means that 65% of the teachers under comparison are below that teacher). Further, an error range is given indicating the uncertainty surrounding the ranks. Table 1 shows an example of the published rankings and margins of error. I have suppressed the names of the teachers in this table. In the *New York Post*, all teachers' and school names were published.

Name	Subject	Multiyear	(Range)	2009-10	(Range)
Teacher 1	Reading	66	(33-86)	50	(11-83)
Teacher 2	Maths	92	(83-96)	80	(60-91)
Teacher 3	Reading	50	(16-80)	43	(9-80)
Teacher 4	Maths	75	(62-86)	65	(42-83)
Teacher 5	Maths	40	(22-58)	78	(55-90)
Teacher 6	Reading	93	(66-97)	96	(84-99)

Table 1: An excerpt from *New York Post* teacher rankings

A glance shows that the margins of error ("ranges") are rather large. The Post stated that the ranks had an average error range of 35 in maths and 53 in English, varying by teacher. For Teacher 3, the percentile rank could be somewhere between 16 and 80, rendering the estimated rank of 50 meaningless. If teacher performance is tied to performance pay, one wonders whether this teacher should get an increment; there is about an equal chance that this teacher performs well above or well below average. Given the large magnitude of error margins, any rankings of teachers based on student achievement must be accompanied by an error range to provide the degree of reliability of the estimates. As can be expected, these error margins will likely be quite large.

Reasons for the unreliability of test score gains as individual teacher performance measures:

Class scores depend very much on a teacher's particular group of students. The large variation in students' academic abilities in a class, together with the unreliability in test scores (since tests are typically quite short), results in random fluctuations of a teacher's class test scores, and contributes to the misidentification of teachers as ineffective or effective.

When we have many years of student data (eg more than 10 years), the error rate of misidentifying a teacher as effective or ineffective will be lower. However, even if student achievement is measured reliably, there are validity issues, as we discuss below.

Validity Issues

Making inferences:

To link student test scores to teacher performance, an inference must be made, since we haven't directly observed teachers teaching (performance). Inferences are conjectures made by people, not proven by statistics. For example, when we see a person driving an old and battered car we might make an inference that the person is not that well-off. This inference may be correct some of the time, or even most of the time, but there are always exceptions. So inferences can never be used as proofs. There is always a margin of error when inferences are made.

In the case of using student test scores to judge teacher performance, the inference has a huge margin of error, simply because so many factors affect student test scores and gain scores (value-added scores). Even if we control for socio-economic status (SES), many other factors have a large effect on the academic growth of students, such as parental support, natural academic ability, motivation, interests, personality, and cultural and ethnic values. Above all, there are always many exceptions where individual students with high SES background perform poorly and students with low SES background perform well, due to factors completely unrelated to school and teachers.

In the case of the *Post*, in a few cases the paper claimed that there was no margin of error. This is likely the result of a misunderstanding. Let's suppose a student has a learning difficulty and therefore has very low gain scores. The margin of error surrounding the student's score may be very small, since we have reliably measured that the student could do little in academic tests. But the margin of error in inferring teacher performance based on this test score is large, in fact, total, since the low gain score has nothing to do with teacher performance.

Example of validity issues:

To illustrate validity issues, we once again use a *New York Post* example. Table 2 shows the rankings of two teachers in a school, in maths and reading.

Name	Subject	Multiyear	(Range)	2009-10	(Range)
Teacher 1	Reading			9	(0-46)
Teacher 1	Maths			90	(67-97)
Teacher 2	Reading			19	(2-62)
Teacher 2	Maths			95	(83-99)

Table 2: Rankings for two Grade 4 teachers in the same school

At first glance it appears the teachers are no good at teaching reading, but extremely good at teaching maths. Or could there be more a reasonable explanation? Are the students mostly of migrant background, excelling in maths but having difficulties with English? Without further information we do not know why they score poorly in reading, but we should at least query whether the rankings really reflect teachers' efforts. This is a validity issue.

More generally, unless students are randomly assigned to schools and classes, class performance is not directly comparable across different schools and different classes. For example, many schools take local students, so the characteristics of a geographical location play a role in student performance, such as schools with more migrant children. Within a school, students may not be randomly assigned to classes. For example, there may be a remedial class or an accelerated class. When there is a non-random assignment of students to classes, class results should not be compared.

Using gain scores instead of test scores does not solve this problem and provide a level playing field for teachers. Higher performing students must have had a faster learning rate than lower performing students, since no-one had reading and maths abilities at birth. Consequently, higher performing students are likely to gain more than lower performing students.

For these reasons, we need to question whether the academic performance of a class actually reflects the efforts of a teacher.

Conclusions

The desire to find quantitative data for making decisions is understandable, but not all data provide hard evidence. In the case of using student performance data for measuring teacher performance, two issues are discussed: reliability and validity. Since test scores are typically based on a once-a-year short test, and there is a wide range of student abilities, the possible variation in class gain scores due to random chance is large in comparison to the teacher effect that we want to measure. As to validity, since we have not directly measured teacher

performance, student outcome can at best be used as indications for further investigation, not as evidence of teacher performance.

From both the theoretical basis and the empirical data now emerging, it is quite clear that student performance data fall far short of the purpose for measuring teacher performance. While we acknowledge that there are differences between teachers in their contributions to student growth, student test data do not provide the power to finely separate effective and ineffective teachers. Attaching high-stake consequences to student performance data can only damage the teaching profession and education more generally.

References

- Darling-Hammond L, Amrein-Beardsley A, Haertel E & Rothstein J (2011). Evaluating teacher evaluation: what we know about value-added models and other methods. *Phi Delta Kappan* 93(6), 8–15.
- Hattie JA (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London & New York: Routledge.
- Leigh A (2010). Estimating teacher effectiveness from two-year changes in students' test scores. *Economics of Education Review*, 29(3), 480–488.
- McCaffrey DF, Koretz D, Lockwood JR, Louis T, & Hamilton LS (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67–101.
- Schochet P, & Chiang HS (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, July 2010, NCEE. Retrieved 2010/10/26, <http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf>
- Wu, ML (2010). Measurement, sampling and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.

A cautionary tale

Governor Cuomo and the effort to destroy public education in New York

David Hursh

As a long-time activist in educational policy, I have observed in New York the continual ratcheting up of high-stakes testing requirements, beginning in the 1990s with the graduation requirement of passing five standardised tests, then, under No Child Left Behind, the requirement for standardised tests in math and reading in grades 3–8 as a means of assessing students, schools and school districts; and finally, with the institution of the Common Core State Standards, requiring standardised tests in every subject not only to assess students, but to determine teacher effectiveness and potentially removing teachers whose students do poorly on the tests (see Hursh, 2007, 2008, 2013).

Furthermore, teachers are increasingly blamed not only for the failings of our education system but also for the increasing economic inequalities in society and the decline of the middle class, a tactic that Michael Apple describes as “exporting the blame” (Apple, 1996).

However, the increasing use of standardised tests to hold accountable and punish students and teachers tells only part of the story. Standardised testing is increasingly used as part of the rationale for privatising education by increasing the number of charter schools — publically funded but privately operated elementary (primary) and secondary schools. Consequently, public education and teachers face the greatest threat yet, one that may mean the demise of public education in New York’s cities and of teaching as a profession.

Governor Andrew Cuomo, a Democrat but not a progressive, recently chaired Camp Philos, a three-day event on educational reform at Whiteface Lodge in the Adirondack Mountains, upstate New York. Many of the invitees were sponsored by a group called Education Reform Now, a non-profit advocacy group that lobbies state and federal public officials to support charter schools, evaluate teachers based on student test scores, and

David Hursh is a professor at the Warner Graduate School of Education at the University of Rochester. His research, writing, and political efforts focus on the rise of neoliberal policies in education. His articles have appeared in numerous journals, including *Journal of Education Policy*, *Discourse: Studies and the Cultural Politics of Education*, and *Educational Studies*. His best-known book is *High-Stakes Testing and the Decline of Teaching and Learning: The Real Crisis in Education*. He also teaches, researches and writes in the area of environmental sustainability, including co-editing a forthcoming special issue of *Environmental Education Research* on environmental education in a neoliberal culture.

eliminate tenure for teachers. Many of the remaining invitees were hedge fund managers, who see charter schools as investment opportunities. Admission to the retreat cost \$1000 per person, an amount teachers can little afford. Even so, when some teachers attempted to register, they were told, "No, thank you."

Cuomo's support for charter schools was made blatantly clear a few months ago when he led a rally at the state capitol promoting charter schools. At the rally he stated: "Education is not about the districts and not about the pensions and not about the unions and not about the lobbyists and not about the PR firms — education is about the students, and the students come first." He then continued to misrepresent the evidence regarding the effectiveness of charter schools, ignoring the fact that charter schools cream off the more capable students, often denying admission to students who are English language learners or students with disabilities. He also seemed to forget that charter schools have more funding per student because they do not have to pay for the space they use in public school buildings, they pay lower salaries to their teachers who are typically young and work under year-to-year contracts, and receive extra funding from corporations and philanthropic foundations who support privatising schooling.

He also forgot to mention that he has received \$400,000 for his upcoming re-election campaign from one charter school operator and another \$400,000 this year from bankers, hedge fund managers, real estate executives, philanthropists and advocacy groups who have flocked to charter schools and other privatisation efforts.

Cuomo often describes New York's schools and teachers as failing. While public schools could do better, especially if teachers were supported in developing culturally appropriate and challenging curriculum, to place all the blame on teachers ignores four major issues. First, test scores are manipulated to yield whatever result current and past commissioners of education desire. As I have detailed elsewhere, results on the standardised tests are entirely unreliable because commissioners have raised and lowered the cut score on tests to portray students as failing or improving, depending on what suited their political interests (Hursh, 2007, 2008, 2013). For example, on the newly instituted Common Core exams, the cut score was set so high as to result in failing 69% of students state-wide and 95% of students in the city of Rochester. Such low passing rates have been used to denigrate public schools and teachers, and as evidence for why education needs to be privatised. Further, because the current commissioner, John King, wants to take credit for improving student learning in the state, he has already guaranteed that the scores on this year's tests will improve, which he can ensure simply by lowering the cut score.

Second, Cuomo and other corporate reformers ignore data that shows that New York's public schools are highly racially and economically segregated; indeed, we have separate and unequal schools. A new study (Kucsera, 2014) by the Civil Rights Project at UCLA confirms what many of us always suspected: New York State has the most segregated schools in the US. Sixty years after *Brown vs the Board of Education* supposedly ended segregation, New York's schools are more segregated than in the past. In 2009, writes Kucsera, "black and Latino students in the state had the highest concentration in intensely-segregated public schools (less than 10% white enrolment), the lowest exposure to white students, and the most uneven distribution with white students across schools" (p1).

Third, Rochester has the fifth highest poverty rate of *all* cities in the US and the second highest of mid-sized cities. Ninety per cent of the students in the Rochester City School District come from families who live in economic poverty. Cuomo, who makes regular announcements on everything from shopping locally for Easter presents to avoiding ticks while hiking, has remained silent on the issue of segregation (Bryant, 2014, April 26).

Fourth, charter schools perform on average no better than publicly administered schools. Since they have the advantage of accepting only the more capable learners, leaving the others behind in public schools, and, in many cases have space provided free by public schools, and receive additional financial support from the Walton Family and other foundations (Rich, 2014), charter schools should have much better results than they do.

Given the weakness of the corporate reformers' arguments, how do we explain their ability to move their agenda forward? First, the corporate reformers aim to control the discourse of public education, portraying themselves and their reform agenda as the only one that aims to improve education for all students, particularly for children living in our urban areas. While Cuomo ignores the more intractable issues of school segregation and child poverty, he claims to support charter schools because "children come first".

In the past he has used observances marking Martin Luther King Jr's birthday to assail teachers as the primary cause for the failures of New York's education system and assert that high-stakes testing responds to King's vision. To be specific, Cuomo claims: "We have to realise that our schools are not an employment program.... It is not about the adults; it is about the children" (Kaplan & Taylor 2012, A-17). Instead, he portrays teachers' unions as special interests and unionised teachers as caring only about their pensions and contracts, while only he and others like him are for the children.

Similarly, he states that "education is not about the lobbyists", portraying himself as above special interests and defying the efforts of lobbyists. Perhaps because Camp Philos brings

together the corporate and political elite who are united in holding teachers and students accountable through standardised tests, ending tenure, decreasing the power of unions, and privatising education, and because most importantly they are not educators, Cuomo imagines them as not the lobbyists they are but merely advocates for equality.

Which leads to the second explanation for the corporate reform success: they have money and lots of it, which not only gives supporters of privatisation access to politicians, such as the Camp Philo retreat, but also supports projects that help them achieve their goals. The Walton Family Foundation, which despises unions and supports charter schools and voucher programs that use public funds to send children to private schools, has since 2000 given approximately \$1 billion to charter schools and their advocates (Rich, 2014, April 25). Likewise, the Bill and Melinda Gates Foundation has poured billions into privatisation efforts and reforms including the Common Core State Standards and exams. On the Common Core alone, research by Jack Hassard, Professor Emeritus at Georgia State, shows “compelling evidence” that Gates has provided \$2.3bn in support of the Common Core, with “more than 1,800 grants to organisations running from teachers unions to state departments of education to political groups like the National Governors’ Association [that] have pushed the Common Core into 45 states, with little transparency and next to no public review” (Schneider, 2014, March 17, p1).

Money buys influence. In March, Gates and David Brooks (2014), *New York Times* editorialist and outspoken supporter of the Common Core, had dinner with 80 US senators. Similarly, the Walton Family Foundation not only funds, according to its own website, one in four charter schools in the US but also funds advocacy groups like Students First, led by Michelle A. Rhee, the former Washington DC schools chancellor who oversaw many of the policy changes funded by Walton.

As Rich (2014) notes: “Students First pushes for the extension of many of those same policies in states across the country, contributing to the campaigns of lawmakers who support the group’s agenda.” The influence of wealthy families such as the Gates and the Waltons confirms the findings of a recent study by Martin Gilens (2013), *Affluence and Influence: Economic Inequality and Political Power in America*, that policy makers enact the preferences of the rich.

All of the above suggests that corporate reformers have used their wealth and power to dominate the education reform agenda and promote the privatisation of public education, increased standardised testing, and the demise of teaching as a profession. Consequently, what hope is there for resisting and reversing the corporate agenda?

In New York and across the country there is increasing resistance to the corporate reform movement as teachers, parents, students, and community members form alliances to combat corporate reforms. Last August, I was one of 12 educators and community members to create the *New York State Allies for Public Education* (www.nysape.org), which offers critical analysis of the corporate reform movement in New York. The number of organisations making up the allies now numbers 50.

Furthermore, critics of corporate reform have influenced the dominant discourse, in particular making economic and racial inequality part of the agenda. For example, critics are using research that reveals the failure to integrate schools 60 years after *Brown vs Board of Education* to make racial inequality an issue. They are also using the 40th anniversary of President Lyndon Johnson's War on Poverty to ask why there is more economic inequality now than at any time since the Great Depression.

And they are using the increasing efforts by Pearson and other corporations to turn schools into profit centres to question the purpose of schooling. Recent hearings held by Commissioner King regarding the implementation of the Common Core curriculum and exams were completely dominated by critics. There have been calls for the resignation of the current commissioner. Lastly, New York State United Teachers organised four hundred teachers to "picket in the pines" at Camp Philo to protest at its exclusion of teachers. The New York State Regents, who make education policy, and the New York State legislature have both acted to implement moratoriums on state initiatives to increase testing of students and teachers. Teachers, parents, and community members are becoming increasing knowledgeable, outspoken and allied regarding the corporate reform movement. The battle is on.

Note

For nine weeks from mid-January to mid-March, I visited teachers, union officials and university faculty in Australia and New Zealand to learn more about the education reform initiatives in both countries. I also gave numerous presentations on the corporate-led education reform movement in the US and, in particular, my home state of New York. You can see my keynote talk to New Zealand primary school teachers and administrators at youtu.be/hW4vZGsLiL4.

References

Apple M (1996). *Cultural politics and education*. New York, NY, US: Teachers College Press.

Dobbin, S. (2013, December 10). New study: Rochester is fifth poorest city in country. *Democrat and Chronicle*. <http://www.democratandchronicle.com/story/news/local/2013/12/10/new-study-rochester-is-fifth-poorest-city-in-country/3950517>

Brooks D (2014, April 18). When the circus descends. *New York Times*. A-23.

- Bryant E (2014, April 26). Governor silent on school segregation. *Democrat and Chronicle*. <http://www.democratandchronicle.com/story/news/local/columnists/bryant/2014/04/26/bryant-governor-silent-school-segregation/8176951>.
- Gilens M (2013) *Affluence and Influence: Economic inequality and political power in America*. Princeton, NJ, US: Princeton University Press.
- Hursh D (2007). "Assessing the impact of No Child Left Behind and other neoliberal reforms in education," *American Educational Research Journal*, 44(3), 493-518.
- Hursh D (2013). Raising the stakes: High-stakes testing and the attack on public education in New York. *Journal of Education Policy*, 28(5), 574-588.
- Hursh D (2008). *High-Stakes Testing and the Decline of Teaching and Learning: The Real Crisis in Education*. Lanham, MD, US: Rowman and Littlefield.
- Kaplan T and Taylor K (2012, January 17). Invoking King, Cuomo and Bloomberg stoke fight on teacher review impasse. *The New York Times*: A-17.
- Rich M (2014, April 25). A Walmart Fortune, Spreading Charter Schools. *New York Times*. A-1. Accessed at <http://nytimes.com/2014/04/26/us/a-walmart-fortune-spreading-charter-schools.html?referrer=>
- Schneider M (2014, March 17). Gates Dined on March 13, 2014, with 80 Senators. <http://deutsch29.wordpress.com/2014/03/17/gates-dined-on-march-13-2014-with-80-senators/>

PISA 2012: Australia's performance

Sue Thomson

Under the umbrella of Australia's national assessment program, all students in Years 3, 5, 7 and 9 complete NAPLAN. Various samples of students also complete a number of international assessments: the Trends in International Mathematics and Science Study (TIMSS) for Year 4 and Year 8 students, the Progress in International Reading Literacy Study (PIRLS) for Year 4 students, and the Programme for International Student Assessment (PISA) which tests 15-year-old students.

PISA has been conducted every three years since 2000, and focuses on reading, mathematics and science. One of these fields is the focus of study in each cycle: reading in 2000 and 2009, mathematics in 2003 and 2012 and science in 2006. A random sample of approximately 15,000 students from about 800 randomly selected schools across Australia were part of the PISA 2012 assessment. As well as completing the assessment, students in PISA also complete a student questionnaire which provides background information such as sex, Indigenous background and language background, as well as students' attitudes and beliefs about school and the particular subject focus. School principals complete a questionnaire which provides information about the school environment.

The results of the PISA assessment are reported in two different ways. The first is the mean score, standardised to a mean across the OECD of 500 and a standard deviation of 100 in the first year of testing. The second is by the percentage of students in ranges of scores termed proficiency levels. This article will focus on PISA reading and mathematical literacy results between 2000 and 2012.

Sue Thomson is the Director of the Educational Monitoring and Research Division at the Australian Council for Educational Research. Dr Thomson's research at ACER has involved extensive analysis of large-scale national and international data sets; she is also involved in several projects involving analysis of the longitudinal data collection associated with the PISA surveys. She has consulted with a variety of government departments at both Commonwealth and state level on a variety of data analysis projects related to TIMSS, PIRLS and PISA.

Australia's performance in reading and mathematical literacy

Figure 1 shows Australia's performance on reading literacy over the five cycles of PISA. The square in the middle of the bar represents the mean, and the bars above and below show the confidence intervals – the scores between which we are 95% confident that the actual mean will lie. When PISA was first conducted in 2000, Australia was outperformed by only one country – Finland. Eight other countries achieved at a level similar to that of Australia.

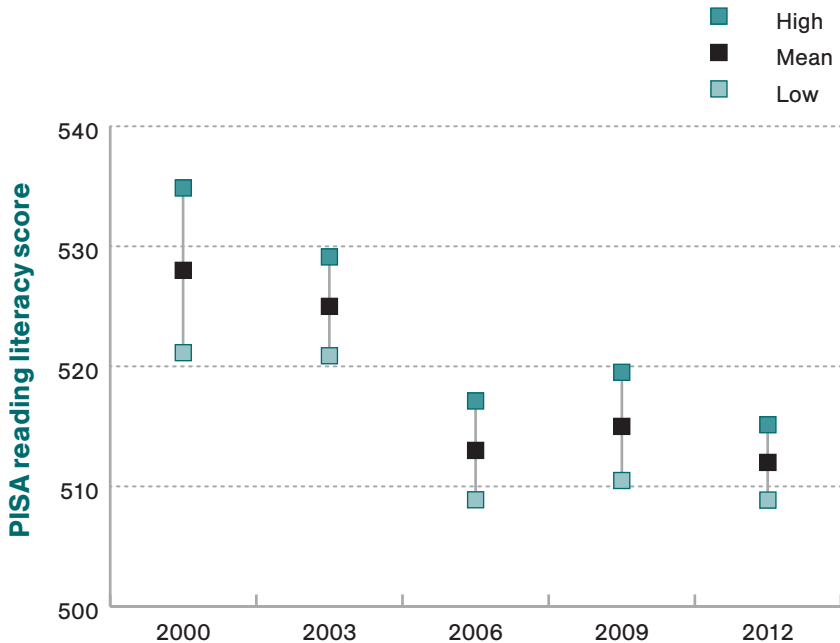


Figure 1. Reading literacy scores for Australia 2000–12

In 2006 the average achievement in reading literacy for Australia declined from 528 to 513, a difference that is statistically significant and substantial in that it represents about half a year of schooling.

In 2009, reading literacy was again the major focus of PISA. Australia's score in reading literacy remained statistically the same as in 2006. Other countries also recorded significant declines in their scores: Ireland, Sweden and the Czech Republic, while seven countries (Chile, Israel, Portugal, Poland, Korea, Hungary and Germany) recorded significant improvements, with gains of between 13 and 40 score points.

Further investigation found that between 2000 and 2009 there was a decline in the proportion of high achievers in reading literacy in Australia, from 17 per cent to 13 per cent, and a slightly higher proportion of low performers, rising from 12 per cent to 14 per cent.

PISA 2003 was the first assessment focussed on mathematical literacy. Figure 2 shows the average scores in mathematical literacy for Australia over the past four cycles of PISA. As with reading literacy, Australia's average score has declined significantly over the past nine years, from 524 to 504 score points. This is again the same as about half a year's schooling.

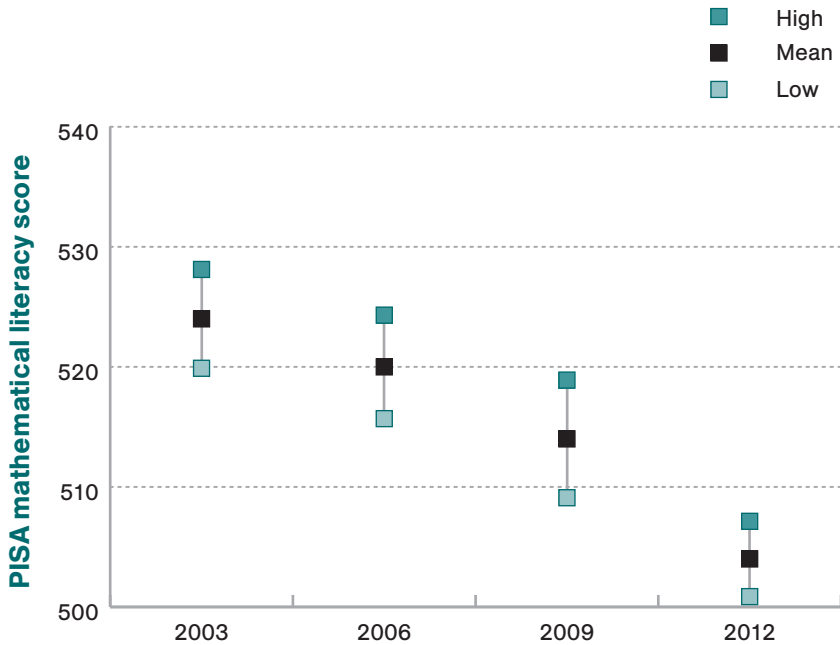


Figure 2. Mathematical literacy scores for Australia 2003–12

While Australia's overall score in mathematical literacy has declined, that of other countries has improved, pushing Australia further down the "ladder". In PISA 2003, Australia was outperformed by only five countries. In 2012 Australia was outperformed by 16 countries, and 12 of these were also in PISA 2003 – either their scores improved or at least stayed the same while Australia's dropped.

How widespread is the decline?

Figure 3 shows the decline in the average reading scores for each state and territory; Figure 4 shows the same for mathematical literacy. Significant differences are shown with an arrow. The two highest achieving states, the ACT and NSW, both had significant declines in scores in reading literacy. Of most concern is that South Australia and Tasmania's score each dropped by 31 score points, which is almost a year of schooling. In all states, the decline was the result of a smaller proportion of students achieving at the highest two proficiency levels.



Figure 3. Reading literacy changes between 2000 and 2009, by state

In mathematical literacy, the decline could be seen in all states other than Victoria (Figure 4).

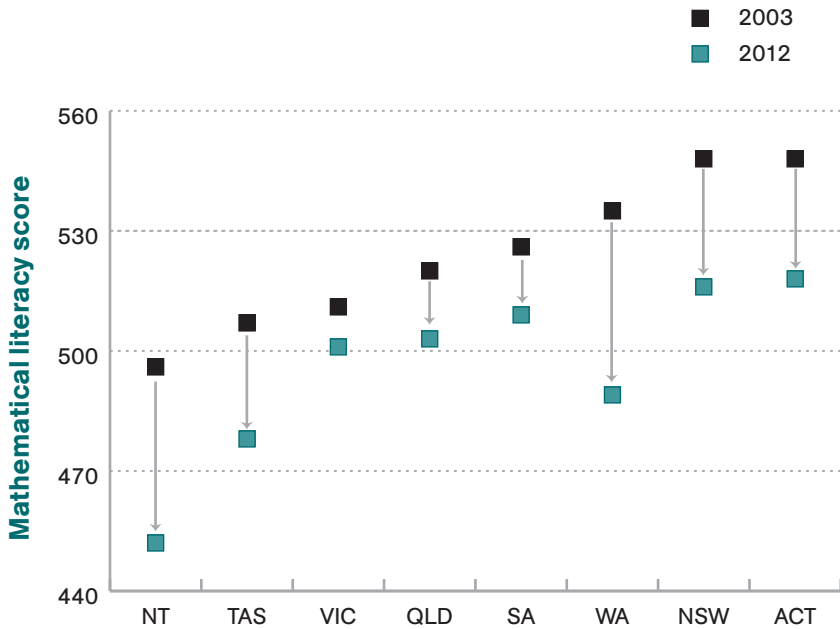


Figure 4. Mathematical literacy changes between 2003 and 2012, by state

Questions of equity

PISA relates achievement scores to a number of student characteristics, including sex, Indigenous status, socioeconomic background, language background, immigrant background and geographic location. Summary data for each of these for reading and mathematical literacy are shown in Table 1.

In reading literacy, scores for males have declined over time, while in mathematical literacy scores for females have significantly declined. In both areas there are significant gender differences, although they are much larger in reading literacy than mathematical literacy.

The gap between Indigenous and non-Indigenous students in both reading and mathematical literacy is large — over 80 score points or more than two years of schooling.

Despite a significant decline in the scores of non-Indigenous students and no decline amongst Indigenous students, the gap has not closed.

The gap that was evident in both reading and mathematical literacy between students from an English-speaking background and those with a language background other than English has, however, closed, with the scores of students with an English-speaking background significantly declining over time.

Similarly, the scores for Australian-born students have significantly declined, whereas the scores for students from an immigrant background have remained statistically the same. As a result of these changes, the scores for students from an immigrant background are significantly higher than those for Australian-born students in both reading literacy and in mathematical literacy.

In both PISA 2000 and PISA 2009 students in metropolitan locations recorded higher average reading literacy scores than students in non-metropolitan locations; over the period there was no significant change in the difference between the two groups. Similarly in mathematical literacy, students in metropolitan areas recorded higher scores than those in non-metropolitan areas, and despite scores for all groups declining, the gap between the two groups in 2012 was similar to that in 2003.

The measure of socio-economic background used in PISA is the measure of Economic, Social and Cultural Status (ESCS). This measure combines information obtained from students about parental education and occupation, and the presence of particular home educational resources, cultural possessions and wealth. The association between both reading and mathematical literacy achievement with the socioeconomic background scale ESCS is shown in the differences in mean scores for the quarters of the distribution of ESCS in Table 1. Those data show that the difference in achievement between the top and bottom quarters in both reading and mathematical literacy is substantial but appears to have reduced a little mainly as a consequence of the drop in the average score of those in the top quarter.

	Reading literacy			Mathematical literacy		
	PISA 2000 Mean	PISA 2009 Mean	Significance of difference between cycles	PISA 2003 Mean	PISA 2012 Mean	Significance of difference between cycles
Sex						
Females	546 (4.7)	533 (2.6)		515 (2.9)	498 (2.0)	*
Males	513 (4.0)	496 (2.9)	*	526 (3.2)	510 (2.4)	
Difference	34 (5.4)	37 (3.1)		-11 (4.3)	-13 (3.0)	
Indigenous background						
Non-Indigenous	531 (3.4)	518 (2.2)	*	526 (2.1)	509 (1.6)	*
Indigenous	448 (5.8)	436 (6.3)		440 (5.4)	423 (4.4)	
Difference	83 (6.7)	82 (6.7)		86 (5.8)	86 (4.5)	
Language background						
English at home	535 (3.6)	518 (2.0)	*	529 (2.0)	506 (1.5)	*
LBOTE	504 (7.5)	509 (8.9)		505 (6.1)	509 (5.2)	
Difference	31 (7.4)	10 (8.3)		24 (6.4)	-5 (4.9)	
Immigrant status						
Australian born	532 (3.6)	515 (2.1)	*	513 (2.2)	501 (1.6)	*
Immigrant background	520 (6.7)	524 (5.8)		526 (6.3)	516 (2.6)	
Difference	12 (6.6)	-10 (5.8)		-13 (6.1)	-15 (2.8)	
Location						
Metropolitan	535 (4.8)	521 (2.9)	*	529 (2.5)	513 (2.0)	*
Non-metropolitan	518 (7.0)	496 (4.0)	*	515 (4.3)	487 (2.3)	*
Difference	17 (8.8)	25 (5.1)		14 (5.1)	26 (3.0)	
Economic, social and cultural status (ESCS)						
Top quarter	587 (4.9)	562 (1.7)	*	572 (2.9)	550 (2.4)	*
Upper quarter	538 (4.5)	532 (1.5)		537 (3.1)	521 (2.7)	*
Lower quarter	516 (3.8)	504 (1.9)	*	513 (2.3)	492 (1.9)	*
Bottom quarter	476 (3.6)	471 (2.1)		479 (4.1)	463 (2.2)	*
Difference (top-bottom)	112 (6.1)	91 (2.7)	*	93 (5.0)	87 (2.9)	

Table 1. Scores and differences on student background characteristics

Equity in learning opportunities

One of the most important indicators of equity is the strength of the relationship between the social background of students and their educational achievement. If this relationship is strong, the educational system is not successful in achieving equitable outcomes, and could be reinforcing educational privilege. In analyses of the variance in scores in mathematical literacy in Australia both within and between schools, it was found that about 20% of this variance in mathematics performance could be explained by either student-level socio-economic background or, much more strongly, the combined pooled effect of student socio-economic background at the school level. However Table 1 shows that there is a wide gap in achievement on PISA between students from a disadvantaged background and students from an affluent background. This difference is in the order of three years of schooling.

Research has identified a number of characteristics that help schools to become “effective” in encouraging high performance from all students, including having adequate facilities and equipment, having well-prepared teachers, providing a safe and orderly environment, and supporting academic success. Figure 5 shows the differences between schools with an average high socio-economic background and schools with an average low socio-economic background on some of the student and school indices that reflect some of these characteristics. These are standardised to a mean over the OECD of 0 and a standard deviation of 1. As it is a little difficult to interpret the size of the difference, effect sizes were also calculated for the difference between the high and low-SES groups. Effect sizes simply provide an indication of the strength of the relationship. Generally effect sizes are referred to as small, medium or large.

The index of quality of school educational resources was derived from six items measuring school principals’ perceptions of potential factors hindering instruction at their school. These factors were: i) shortage or inadequacy of science laboratory equipment; ii) shortage or inadequacy of instructional materials; iii) shortage or inadequacy of computers for instruction; iv) lack or inadequacy of Internet connectivity; v) shortage or inadequacy of computer software for instruction; and vi) shortage or inadequacy of library materials. Higher values on this index indicate better quality of educational resources. The index scores for both high and low SES schools were positive, indicating that most Australian schools were not hindered to a great extent by a lack of educational resources; the effect size for the difference between high and low SES schools was large, with a difference between the two of about three-quarters of a standard deviation.

The index of disciplinary climate, next in Figure 5, was derived from students' reports on how often the following happened in their lessons: i) students don't listen to what the teacher says; ii) there is noise and disorder; iii) the teacher has to wait a long time for the students to quieten down; iv) students cannot work well; and v) students don't start working for a long time after the lesson begins. In this index higher values indicate a better disciplinary climate. Students in high-SES schools reported a better disciplinary climate than was on average over the OECD, while students in low-SES schools reported a substantially lower level than the OECD average. The effect size for this was medium, about 0.4 of a standard deviation.

There was a large difference on the third variable under examination, the index of teacher shortage. This was derived from four items measuring school principals' perceptions of potential factors hindering instruction at their school. These factors are a lack of: i) qualified science teachers; ii) qualified mathematics teachers; iii) qualified English teachers; and iv) qualified teachers of other subjects. Higher values on this index indicate school principals' reports of higher teacher shortage at a school. The index for high SES schools scored about half a standard deviation higher than the OECD average, and almost one full standard deviation higher than low SES schools. It is clear from this that while teacher shortages are not an issue for high SES schools they most certainly are for low SES schools, whose score on this index reflects principals' reports that instruction is hindered to a greater extent than on average across the OECD by teacher shortages in key areas.

The index quality of physical infrastructure was derived from three items measuring school principals' perceptions of potential factors hindering instruction at their school. These factors are: i) shortage or inadequacy of school buildings and grounds; ii) shortage or inadequacy of heating/cooling and lighting systems; and iii) shortage or inadequacy of instructional space (eg classrooms). Higher values on this index indicate better quality of physical infrastructure. As with school resources, both low and high-SES schools reported a better level of physical infrastructure at their schools than on average across the OECD. However there was still a difference with moderate effect size between high and low SES schools in Australia.

The index of student-related factors affecting school climate was derived from school principals' reports on the extent to which the learning of students was hindered by the following factors in their schools: i) student truancy; ii) students skipping classes; iii) students arriving late for school; iv) students not attending compulsory school events (eg sports day) or excursions; v) students lacking respect for teachers; vi) disruption of classes by students; vii) student use of alcohol or illegal drugs; and viii) students intimidating or bullying other students. Higher values on this index indicate a positive student behaviour. With a difference of almost one and a half standard deviations, the difference between high and low SES

schools in Australia is very large. The school climate at high-SES schools is substantially better than at low-SES schools, with the learning of students at low SES schools much more likely to be affected by these factors than those in high SES schools.

Last in Figure 5 is teacher morale, as reported by the principal. That of teachers at high-SES schools is substantially higher than the OECD mean, while the morale of teachers at low-SES schools is significantly lower than the OECD average. It is perhaps not surprising given the results indicated previously.

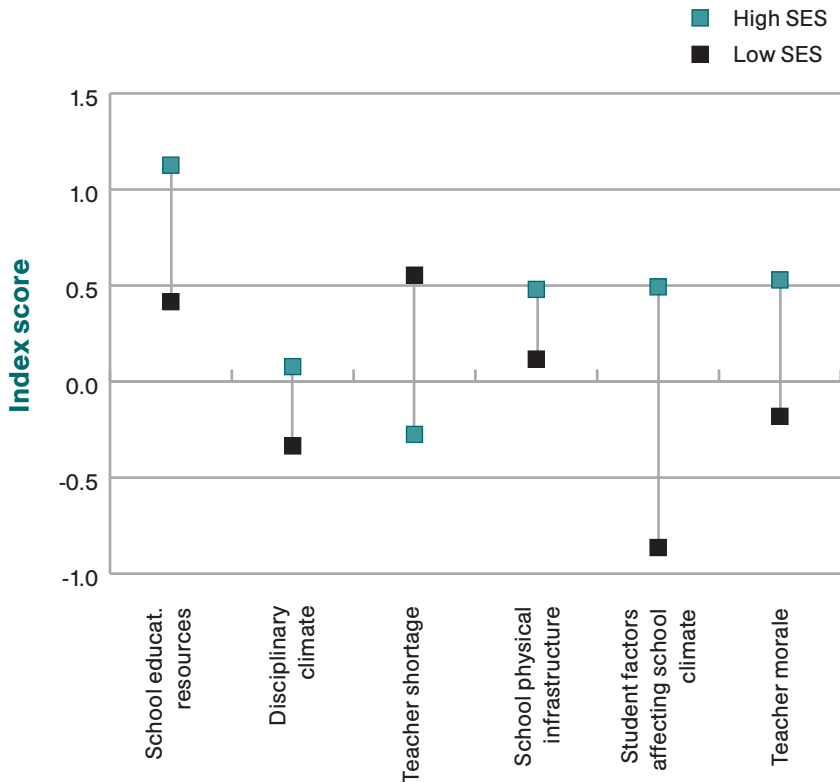


Figure 5. Difference between schools with an average high SES and an average low SES

Conclusion

Australia's score in PISA in both mathematical literacy and reading literacy, while still relatively high in world terms, is slipping, while that of some other countries is staying the same or improving. The change seems to be that a lower proportion of young people are achieving at the high levels and a higher proportion achieving at the lower levels.

Australia has a number of equity issues, with males achieving at a lower level than females in reading literacy but outperforming them in mathematical literacy, Indigenous students scoring at a level about three school years behind their non-Indigenous counterparts, students in regional and remote areas falling behind their counterparts in the cities, and students from poor backgrounds performing at a level almost three years behind their counterparts in high SES schools.

If Australia is to improve its PISA scores, in other words improve educational outcomes for all students, the best way is to improve the performance of students who have fallen behind, and to extend the performance of those who are capable of being extended.

The PISA data can also show where some of the inadequacies of the school system are. By comparing schools with a high average socio-economic background with those with a low average socio-economic background, it is easy to begin to see where some of the differences lie, and perhaps where some of the obstacles lie for all schools to be effective.

The importance of class size

David Zyngier

Class size research has a protracted and controversial history, especially in the USA, England, and Australia. Is there evidence that pupils taught in smaller classes do better in academic and other non-cognitive outcomes than pupils in larger classes?

Many policy-makers and political commentators suggest that funding isn't the problem in Australian education. They claim that much of Australia's increased expenditure on education in the past 20 to 30 years has been "wasted" on efforts to reduce class sizes, arguing that this extra funding does not lead to better academic results.

Most of this policy advice and commentary relies heavily on Jensen's report (2010) on Australian education and teacher quality. Jensen suggests that the majority of studies around the world have shown that class size reductions do not significantly improve student outcomes, and that the funds should have been redirected toward enhancing teacher quality. Although the results of individual studies are always questionable, a range of newer peer-reviewed studies on the effects of small classes have now emerged, and they throw into doubt this advice.

In Australia, commentators and politicians alike point to high-performing systems such as Shanghai, Hong Kong, South Korea, Taiwan and Singapore, where large class sizes are the norm, as evidence that reducing class sizes is a futile exercise. But research indicates that students from Confucian heritage cultures are socialised in ways that make them amenable to work in large classes, so that management problems are minimal and teachers can focus on meaningful learning using whole-class methods. An educational system forms a working whole, each component interacting with all other components. Isolating any one component (such as class size) and transplanting it into a different system shows a deep misunderstanding of how educational systems work.

David Zyngier is a senior lecturer in curriculum and pedagogy in the Faculty of Education at Monash University. He was previously a teacher and school principal. His research on democracy in education was awarded \$366,000 from the Australian Research Council for 2013–16. His research focuses on teacher pedagogies that engage all students but in particular how these can improve outcomes for students from communities of disadvantage, focusing on issues of social justice and social inclusion.

Reducing class size in order to increase student achievement is an approach that has been tried, debated, and analysed for many decades. The premise seems logical: with fewer students to teach, teachers should achieve better academic outcomes for all students. For those who choose private education for their children in Australia, it is often cited as a major consideration. However, for policymakers there are three major questions to answer with the adoption of any change or new program: how effective will the change be; how much will it cost; and what are the problems of implementation, including the support or opposition of the stakeholders — in this case principals, teachers and parents — and those who implement it?

An accurate determination of actual class sizes in Australia is problematic. Moreover, different states and territories collect data on class sizes at different times of the year; students and teachers come and go; and teaching groups change. Student-teacher ratios (STRs) are calculated by dividing the full-time equivalent students on a school's roll by the full-time equivalent number of qualified teachers. STRs are different from class sizes because they also count teachers who are not at the "chalk-face", such as library, welfare, careers teachers and principals. All the enrolled students are divided by all the teachers in the school, yet it should not be assumed that teachers entered into the ratio are teaching for all of the time. Past research has too often conflated STR with class size.

In 2010 Australia's average public primary class size (not STR) was 23.2 — above the OECD average of 21.3 and EU average of 20. This compares to 15 in Korea; 17 in Germany and the Russian Federation; 19 in Finland; 20 in the UK, Poland and Luxembourg; and 26 in India (OECD 2013). Class sizes are also smaller in both the independent and Catholic sectors in Australia.

Policy-makers, politicians and media too often discuss data about class sizes and their impact on student learning without an evidence base, relying largely on second-hand research or anecdotes. Too frequently, advocates for particular positions select their evidence, conveniently ignoring research that raises questions about their favoured position.

Advocates for and against class-size reduction have engaged in or been accused of engaging in such cherry picking for as long as there has been research on this issue. (Whitehurst and Chingos 2011, 3)

I found only two authors during my research study who supported the notion that smaller class sizes do not produce better outcomes to justify the expenditure on them. Hanushek and Hoxby seem to stand alone in their findings that class size reduction has little or no impact on student academic outcomes — yet they are disproportionately referred to for

evidence here in Australia. In a 2011 court case in the USA about school funding, the judge commented on Hanushek's evidence submitted to that trial:

Dr Hanushek's analysis that there is not much relationship in Colorado between spending and achievement contradicts testimony and documentary evidence from dozens of well-respected educators in the state, defies logic, and is statistically flawed. ... The data underlying Dr Hanushek's opinions [are] questionable or problematic and I found him to lack credibility.

Education researchers have refuted the work of Hanushek and Hoxby. They point out that Hanushek (and Jensen) do not examine class size directly, but rather through a proxy measure intended to represent it (student-teacher ratio). While teacher quality (and the quality of teacher preparation) is at the heart of the effectiveness of almost any reform, conflating STR with class size reduction fails to focus on the mechanisms thought to be at work in smaller classes.

Hanushek has not responded well to such criticisms; rather, he has found reasons to quarrel with their details and to continue publishing reviews, based on methods that others find questionable, claiming that the level of school funding and the things those funds can buy, such as smaller classes, have few discernible effects. Political conservatives have extolled his conclusions, complimented his efforts, and asked him to testify in various forums where class-size issues are debated. And in return, Hanushek has embedded his conclusion about the lack of class-size effects in a broader endorsement of a conservative educational agenda.

The highly selective evidence being used to support current advice to state and federal education ministers is based on flawed research. The class size debate should now be more about weighing the cost-benefit of class size reductions, and how best to achieve the desired outcome of improved academic achievement for all children regardless of their background. Further analysis of the cost-benefit of targeted reductions is therefore essential.

Many credible and peer reviewed research projects have concluded that:

- Extra gains found for long-term attendance in small classes in the early grades continue to appear when students returned to standard classes in the upper grades
- These gains in the early grades appear not only for tests of measured achievement, but also for other measures of success in education
- The greater gains experienced by students from traditionally disadvantaged groups are retained when those students returned to standard classes

- The extra gains from small classes in the early grades are greater the longer that students are exposed to those classes
- Extra gains from small classes in the early grades are larger when class size is reduced to fewer than 20 students
- Evidence for the possible advantages of small classes in the upper grades and high school is so far inconclusive.

Reducing class sizes or adding extra teachers requires a new approach to teaching — without adequate professional development, the innovative 21st century teaching spaces provided as part of the Building the Education Revolution can do more harm than good. As Hattie explains, the problem is that teachers in smaller classes are adopting the same teaching methods as in their previously larger classes. Many of the more powerful influences Hattie identifies clearly show that teachers would be even more effective with smaller classes.

Class size reduction and equity

It is evident that for certain groups of children (Indigenous, low SES and culturally, linguistically and economically disenfranchised (CLEDE) students in the early years, and children with learning and behavioural difficulties), smaller class sizes and increased STRs are very beneficial.

This holds for student learning outcomes, behavioural modification, and teacher satisfaction. As Lamb, Teese and Polesel have shown, with the increasing residualisation of public schools caused by the flight of cultural capital — itself a result of years of federal and state neglect and artificial choice programs promoting private schools — public schools have a larger proportion of problematic learners, disadvantaged and refugee families, and students at risk of school failure, but have larger class sizes than ever before in comparison with most private schools.

Class size reduction is about equity — any policy debate must start with the basic inequality of schooling, and aim to ameliorate the damage that poverty, violence, inadequate childcare and other factors do to our children's learning outcomes. It must look at the strategies, pedagogies and practices that could mediate those differences, and "the investments that we are willing to make as a society to put success in reach of all children" (Graue et al. 2005, 31).

If smaller classes are introduced in the current policy context of high-stakes testing, together with the inadequate funding highlighted by the Gonski Review, we can expect

minimal achievement outcomes. Additional resources would acknowledge the deep-seated inequities at the core of Australian schooling, but would not be enough.

Class size reduction is part of a system of reforms and problems that need to be considered in a coordinated manner, in relation to both the practice and research of schooling. It necessitates implementation that "connects the utilisation of the resources for class size reduction with all curricular, administrative, and institutional efforts that shape teaching and learning" (Graue et al. 2005, 32).

Recommendations for policy change

The strongest hypothesis about why small classes work concerns students' classroom behaviour. Evidence is mounting that students in small classes are more engaged in learning activities, and exhibit less disruptive behaviour.

The following policy recommendations and principles are therefore suggested:

- Class size is an important determinant of student outcomes, and one that can be directly determined by policy. Any attempts to increase class sizes will harm student outcomes. (Schanzenbach 2014)
- The evidence suggests that increasing class size will harm not only children's academic results in the short run, but also their long-term success at school and beyond. Money saved by not decreasing class sizes may result in substantial social and educational costs in the future. (Schanzenbach 2014)
- The impact of class-size reduction is greater for low-income and minority children
- While lower class size has a demonstrable cost, it may prove the more cost-effective policy overall in closing the widening gap between the lowest and highest achievers, even in tight budgetary conditions
- Professional development for all staff involved will increase their knowledge of, and readiness to use, techniques that are particularly suited to small class environments
- Specific classes and specific year levels should be targeted for reduced class sizes
- Further research is needed into the exact cost of targeted class size reductions for CLED communities and other disadvantaged learners
- Further research is needed into the specific teacher pedagogies that are more appropriate for smaller classes.

Schools should look for ways to produce the small class size effect by lowering class size specifically for certain periods of instruction in numeracy and literacy classes. If class size could be reduced just for these lessons, by redeploying existing staff with the addition of special literacy and numeracy teachers, it would be theoretically possible to have small

classes (average of 15 pupils) at a much lower additional cost. This approach is used by some principals to deliver smaller class sizes in literacy and numeracy but is not yet general practice for disadvantaged groups and learners with higher needs. Targeted class size reductions combined with other proven methods of improving achievement would be a more cost-effective means of increasing student achievement.

Writing about the USA, Haimson (2009) concludes:

Many of the individuals who are driving education policy in this country ... sent their own children to abundantly financed private schools where class sizes were 16 or less, and yet continue to insist that resources, equitable funding, and class size don't matter – when all the evidence points to the contrary.

The same might be said of Australia.

Note

This is a summary of a major review published in April 2014 and available free online in the ANZSOG Journal *Evidence Base*: bit.ly/1qutgxJ (PDF). All references are to be found there.

Interview: Pasi Sahlberg

On lessons from Finland

Interview by John Graham

JG Finland is seen to have one of the best schooling systems in the world. What elements of the Finnish system do you think make the difference and elevate the performance of its students above those in many other countries?

PS Finland may be seen as having the best school system in the world by foreign media and some others but certainly not by many Finns. When the OECD released its first PISA results in 2001, it struck many by surprise. Finns were among those, because Finland had never intended to be high in the PISA league tables. The way education is seen in Finland – among educators and citizens alike – is very different to how it is seen in many other places where nations compete against each other to see who will be the best. For the Finns, what matters in education is that all children have opportunities to succeed and that each of them feels happy and well in school.

Having said that, there are some visible elements of the Finnish system that do make a difference. I would just mention three here. First, we are strong believers in public schooling and therefore carefully manage licences to operate any other types of schools. There are about 75 independent but publicly funded schools in Finland (such as Steiner schools, religious schools and university teacher training schools) but they are part of the public school network. We do everything we can to encourage parents to choose their own neighbourhood school for their children within the free choice they have regarding where their children will learn.

Pasi Sahlberg is a Finnish educator and scholar. He worked as a schoolteacher, teacher educator and policy advisor in Finland and has studied education systems and reforms around the world. His expertise includes international educational change, the future of schooling, and innovation in teaching and learning. His best-selling book *Finnish Lessons: What can the world learn from educational change in Finland* (Teachers College Press, 2011) won the 2013 Grawemeyer Award.

He is a former Director General of the Centre for International Mobility and Cooperation in Helsinki and currently a visiting Professor of Practice at Harvard University's Graduate School of Education.

Second, we have embedded a comprehensive, early intervention system to identify and support those children who have any kind of special need in school. Every school must have what we call a "pupil welfare team" that is responsible for making sure that all students are properly supported and helped in every school. Special education services currently include about one third of all basic school (grades 1 to 9) pupils. As a result, grade repetition is very low and graduation rates from basic school are therefore close to 100 per cent.

Third, in order to be successful with the earlier two elements, Finland has a particular system of selecting and educating its teachers. In the late 1970s the Finnish Government decided to elevate the teaching profession to the same level as that of other highly-valued professions, such as medicine and law, by making an academic masters degree the basic teaching qualification. Teacher education in Finland is based on research and this has enabled us to enhance the public trust in teachers and their professional responsibilities. Teachers in Finland enjoy social prestige, and teaching is seen as a competitive career choice by many young Finns. Finnish schools are truly professional learning communities with a considerable amount of autonomy and freedom to find the best ways to support pupils' learning.

JG One of the newsworthy items from the PISA 2012 testing program was the slight decline in Finland's performance compared to previous years. How did the media and politicians in Finland react to this decline? Do you have any explanation for Finland's dip in performance in 2012?

PS There is a systematic way of monitoring educational performance in Finland through national evaluations and research. PISA is, in a way, a tool to add value or confirm national findings concerning system performance. Therefore our own domestic data had already indicated that there has been a slight overall decline since the mid-2000s in reading and mathematics learning and, to a lesser degree, in science. Our authorities indicated far before PISA 2012 became public, in December 2013, that Finland was not likely to perform as favourably in PISA this time. The media reported PISA 2012 results as it always does by publishing the international league tables concluding that Finland is no longer the world leader. The reaction of politicians was that Finland is still #2 in the OECD family and therefore doing well.

But the question of why learning outcomes in reading and mathematics in Finland are declining is an important one. One part of the explanation is that there are so many other education systems in the OECD that have adjusted their education policies and funding to enhance their PISA scores. It has almost become a norm to set your

national targets so that your country is among the top five PISA performers in the future. This is not what the Finns have done. The focus in Finnish schools has actually shifted towards arts, social sciences and creativity rather than increasing attention to reading, mathematics and science.

One persuasive argument used to explain declines in student learning in Finland holds that much of that negative trend is associated with the performance of Finnish boys in school. My own data suggests that if Finnish boys were performing at a similar level to Finnish girls (as they do in other OECD countries), there would be no change in Finland's performance in PISA. Furthermore, if Finnish boys did as well as girls in reading, mathematics and science Finland would perform close to Singapore's level. So, it looks like we have a specific challenge to make learning more inspiring for our teenage boys.

JG More generally on PISA, the effect of PISA rankings has a profound and increasing impact on the education landscape in Australia and many other countries. The decline in performance of Australian 15-year-olds in PISA 2012 has been used by governments around the country to justify their favourite education "reform" packages. I also note the recent letter sent by a worldwide list of education academics and others to the OECD deploring the negative consequences of PISA on schooling. What is your view about PISA, its growing influence and its use by governments to justify reforms?

PS Australia should be careful not to make too many quick conclusions about its PISA performance. A number of new "competitors" have joined PISA since its inauguration in 2000 so the nature of game has changed. Some of these newcomers — Singapore, Macao, Hong Kong and Shanghai — have taken the top positions in PISA league tables. This obviously affects where Australia or Finland will appear in these global rankings. We should also know that all of the above mentioned top-ranking education systems have extensive and very expensive after-school tutoring systems that most pupils must attend in order to fulfil their parents' expectations.

Your question about the value of PISA is like asking what do you think about fire! They are both useful and can benefit our lives significantly if we know how to deal with them. Unfortunately PISA is often like a box of matches in the hands of a child. PISA certainly has had negative consequences in some places where it has taken the driver's seat in determining priorities in national education policies. There are a number of countries now (including Australia) that have formulated their goals in education to be on the top of the global league tables. An over-reliance on reaching such targets, by insisting that schools and teachers focus on a narrow area of academic achievement at the expense of broader learning and personal development goals, may have worrying effects later on.

On the other hand, one could claim that without PISA the global geography of education would look very different. I am afraid that we would see even more market-based solutions and privatisation of public education than is currently happening. Certainly we would not be speaking about the key role that equity has in building successful education systems as we do today. And, most concretely, I wouldn't be giving this interview for Australian teachers to read if there was no PISA.

JG One of the outcomes of PISA has been the elevation of certain education systems – Finland, Shanghai-China, Korea etc – to be exemplars for other countries to try to emulate. What is your view about this? Can education systems be transported from one country to another?

PS It is actually very unfortunate that despite the words of warning by the OECD itself and many educators, including me, there are so many who desire to imitate the most successful education systems in PISA in the hope of finding solutions to their own system's challenges. I have been very disappointed by how poorly people in general understand what PISA is actually able to reveal. Most people, educators included, seem to perceive PISA as a global league table that is like a thermometer showing how good or poor the health of your school system is.

But the OECD is very clear about what a successful education system is. It must have high student achievement in all measured domains, it must have a high level of equity that suggests students' socio-economic background is not a strong indicator of their achievement, it must have high overall levels of participation in education (including high graduation rates), and it must ensure both human and financial resources are used efficiently to accomplish these results.

There is a lot we can all learn about these aspects of "high performance", but you should not think that, by redesigning your own education system according to the three elements that have driven Finland's success, things would get any better. My own theory at the moment is that in all successful education systems there are cultural, economic and social factors external to schools which act as powerful drivers of high educational performance equal to those found within school systems. My next book, *Invisible Lessons*, is about these hidden elements behind successful educational performance.

JG You have been responsible for conceptualising and naming a set of education reforms, which have become central to school systems in Anglo-Saxon countries such as the US, the UK and Australia, as the Global Education Reform Movement or GERM. What

is GERM? Does it work? Why do you think governments and education authorities in these countries are so determined to implement GERM policies?

PS Well, GERM is an unofficial education policy orthodoxy that many formal institutions, corporations and governments have adopted as their official program in educational development. This global movement includes some welcome elements that have strengthened the focus on learning, encouraged access to education for all, and emphasised the acquisition of knowledge and skills that are relevant in the real world. But GERM also has symptoms that indicate it may be harmful to its host: driving education reforms by competition, standardisation, test-based accountability, fast-track pathways into teaching and privatisation of public education. My own view is that GERM is a real phenomenon and it has been successful in finding hosts all over the world.

The reasons for the global prevalence of GERM include its common sense logic, examples set by some Anglo-Saxon countries, and the increasing presence of private corporations in school improvement. What has been the effect of GERM so far? PISA, dating from the year 2000, clearly shows that none of the GERM-infected school systems — England, the US, Australia, New Zealand, Canada, the Netherlands or Sweden — has been able to improve educational performance, contrary to the policy promises made when these GERM solutions were chosen to be centrepieces in national education reform programs.

JG *Australia has one of the most privatised school systems in the world. There is now a push to break up the state public systems of schooling by introducing the equivalent of charter schools or the English academies. What is your view about the role of public education systems? Do you think the charter school idea (presently being called Independent Public Schools by our Federal Government) is a good one?*

PS Privatisation of public education is one of the central goals of GERM. Research shows how efforts to privatise education by vouchers or alternative governance models have not brought promised improvements of learning or efficiency gains. This is true for charter schools in the US, Swedish “free schools” and the Chilean experiment. The OECD has a clear message to all those who hope to enhance their education systems by market-based solutions. In its 2012 report, *Equity and Quality in Education*, the OECD concludes: “School choice advocates often argue that the introduction of market mechanisms allows equal access to high quality schooling for all. However, evidence does not support these perceptions, as choice and associated market mechanisms can enhance segregation. The highest-performing education systems across the OECD countries are those that combine quality with equity.”

I think school choice as a way of improving entire education systems is more myth than fact. The question is not, however, choice or no choice. It is about whether we have a good school for all children or just for some. In the end we need to work out how we manage parental choice so that it doesn't harm equity.

JG There is a current obsession with the need to improve "teacher quality". There are a series of initiatives to set more stringent standards for entry into the profession and have more rigorous teacher performance evaluation processes. What is your perspective on these teacher quality arguments?

PS As I see it, another myth is that teaching is easy. In other words, anyone can teach if you are smart and interested in spending time with people. What Finland and Singapore have done, for example, is recognise that teaching is a difficult profession. It requires complex knowledge and skills similar to those that medical doctors and lawyers use in their work. I am all for raising the standards of entry into the teaching profession so that it is on a par with other highly valued professions. The introduction of strict quality controls at the entry point into teaching make more rigorous teacher evaluation processes redundant, as is the situation in Finland and Singapore.

Another common myth of teacher quality is that "the quality of an education system cannot exceed the quality of its teachers". Those who believe this tend to see teaching as an individual craft where the individual teacher is disconnected from other teachers and the professional community they share. But to me at least, school teaching is a team play where leadership, a shared dream and common professional qualifications help the team to be more than its individual players. That is one reason why leadership in schools has become such a valuable currency.

JG Can you explain the nature and level of professionalism that teachers have in Finnish schools?

PS Teacher professionalism has four manifestations in Finnish schools. First, teachers have the main role in planning and designing their curricula. This autonomy and responsibility for planning is considered a cornerstone of teacher leadership and professionalism.

Second, teachers are free to choose the best possible teaching methods to accomplish the learning goals described in the school curriculum. Third, teachers are responsible for assessing and grading their students according to the guidelines stipulated in their school curriculum. Finally, professionalism in Finnish schools requires

that teachers are active members in their professional communities. Academic research-based teacher education in Finland focuses on these four aspects of professionalism. School leadership is then the glue that brings these elements of teacher professionalism together for the good of the school.

JG What does Finland do to develop effective school leadership? How do you become a principal in a Finnish school? Are school principals generally satisfied with their role?

PS There is a strict rule in Finland that school principals must be qualified to teach in the same schools that they lead. This means that a high school mathematics teacher may not be appointed as a primary school principal without a primary school teacher's qualification. Effective school leadership starts there. In Finland we expect that school principals must always be experienced teachers who have the personal characteristics to lead other teachers and a school. Before appointment to school leadership posts, successful candidates must have a set amount of leadership training typically offered by Finnish universities, a positive track record as a teacher and a suitable personality.

School principals are now often appointed for a fixed term (five to seven years). Increasing bureaucracy and a tightening economic situation has affected principals' working conditions as well. There are more experienced teachers who would have earlier looked for leadership posts in schools who are now deciding to remain in their teaching jobs — often due to the reasons mentioned above.

JG Is a commitment to equity highly valued in the Finnish education system? What sort of policies (eg funding) and procedures are used to ensure equitable outcomes for Finnish students?

PS Equity is the foundation of Finland's education system. Since the early 1970s Finland's education policies have aimed at building a system that provides all students with equal educational opportunities to succeed in school regardless of their family background or domicile. Today Finland is one of those successful education systems that has high learning outcomes and system-wide equity.

There are certain principles that Finland has followed in creating its equitable school system. First of all there is equitable school funding that disburses resources to schools based on their real needs. This is pretty close to what the Gonski Report suggested for Australia. Secondly, there is a universal and well-resourced special education system that is able to provide support to all those in need early on and without being labelled in school.

Thirdly, there is a systematic way of embedding pupil health and wellbeing support on a daily basis as part of the work of every school. This includes healthy school meals, health checks, dental checks and counselling for all children. Finally, Finland has a balanced curriculum that is based on a realisation that multiple intelligences exist in every classroom. Equity is enhanced when all students have access to high quality arts, music, physical education and other non-academic subjects in the same way that they study reading, mathematics, science and other academic subjects.

Most Finns believe that a strong public education system is the best way to maintain and enhance both quality and equity in education. The international evidence is also clear: the most successful education systems are those that combine quality and equity in their education priorities, and that cultivate education as a basic human right through public service for all children.

Professional Voice

Subscription form and tax invoice

ABN: 44 673 398 674 | All prices are inclusive of GST

PO Box 363, Abbotsford 3067 | Tel: 03 9417 2822 | Fax: 1300 568 078 | Web: www.aeuvic.asn.au

Professional Voice is published tri-annually and aims to stimulate debate within the Australian education community. Covering issues of curriculum, pedagogy, leadership, educational practice and related topics, *PV* is positioned to build awareness of innovation, best practice and reform within educational settings from early childhood to TAFE, adult provision and disability centres. *PV* is published by the AEU Victorian Branch and is available free to financial members by emailing aeunews@aeuvic.asn.au.



Name:

Address:

..... Postcode:

Daytime telephone number:

Email:

Please subscribe me to *Professional Voice* beginning with issue:

Please send me the following back issue(s) of *Professional Voice*

Current issue \$10 | back issues \$10 | one year (3 issues) \$25 | overseas (airmail) +\$5 per issue

Total amount payable: AUD\$ (includes GST)

Please find enclosed my cheque: (in AUD\$ payable to the Australian Education Union Victorian Branch)

Please debit my credit card: Visa Mastercard Bankcard

Card Number: Expiry:

Cardholder's Name:

Signature: Date:



Please use a separate form for each subscription.

Professional Voice

Back issues

The following back issues are still available, free to AEU members from aeunews@aeuvic.asn.au.

Non-members and institutions can order them for \$10 each using the form overleaf. These and other issues can be found online at www.aeuvic.asn.au/pv.

PV 9.3: Global Education Reform Movement

With an editorial overview of the GERM agenda, stories include a look at NAPLAN and assessment, collaborative teaching, class sizes and the models of reform being pursued in America and the UK.

PV 9.1: Equity and Disadvantage

The autumn 2012 edition takes a broad look at equity issues. Alan Reid argues that governments have fallen in love with quick fixes and easy answers. Alan Smithers notes that choice and standards policies in England failed to increase equity because they were not part of a well-designed system. Tony Vinson says that investment in early years is not only the best investment we can make in society, it is a moral imperative.

PV 8.3: The National Agenda

This edition of PV casts a critical eye at the national schools agenda, from funding to curriculum to the new teacher standards. A strong line-up of writers includes Chris Bonnor, Alan Reid, and Lyndsey Connors. Also in this issue, part one of a major interview with leading US academic Linda Darling-Hammond.

PV 8.2: Partnerships with Parents

This edition of Professional Voice examines the relationship between schools and families - the ways and reasons why parents and carers are involved in schools and their role in teaching and learning.

PV 8.1: The NAPLAN Debate

The genesis of this Winter 2010 edition of Professional Voice was a national symposium held in Sydney in July.

Its title was "Advice for Ministers and ACARA on NAPLAN, the use of student data, My School and league tables". Contributors include Alan Reid, Margaret Wu, Allan Luke and Brian Caldwell.

PV 7.2: Beyond Edu-Babble

With education policy paralysed by management speak and business cliches, Guy Claxton, Howard Gardner, Robin Alexander and others attempt to cut through the cant with fresh thinking on the challenges facing education.

PV 6.2: Early Years Education

The second of our three-issue survey looks at developments including the new early years framework, effective literacy programs and the national reform agenda, plus analyses of early intervention and phonics programs.

PV 6.1: Post-compulsory Education

This is the first of three issues each looking at different phases of the education continuum. Articles look at developments and issues in TAFE, higher education, technical and vocational education and training.

PV 4.1: Teacher Quality and Quality Teaching

Includes Andy Hargreaves on teaching in the knowledge society, and the implications for schools of globalisation; Leonie Rowan and Chris Bigum on the challenges of measuring quality in teaching and education; and Lyndsay Connors on the part that class still plays in education.

Testing Times



Editorial: Testing times

John Graham

Flaws in the lens

Alan Reid

Linking student test scores to teacher performance

Margaret Wu

A cautionary tale

David Hursh

PISA 2012: Australia's performance

Sue Thomson

The importance of class size

David Zyngier

Pasi Sahlberg on lessons from Finland

Interview by John Graham