
12 Australian perspectives on benchmarking: the case of school education*

Peter Dawkins¹
Victoria University

12.1 Introduction

Improving the educational outcomes for children and young people is central to the nation's social and economic prosperity. In Australia, there are a complex set of Commonwealth–State relationships and division of powers in relation to school education (see Banks, Fenna and McDonald, this volume). This chapter focuses on school education in the Australian federal context and the major benchmarking developments that have occurred in this area over the last twenty years.

The Australian States have constitutional responsibility for school education, including the administration of government schools; development and delivery of curricula; and the regulatory conditions to ensure quality standards across all schools (including non-government schools).

Although State governments have primary responsibility for education, the Commonwealth has assumed an increasingly important role, driven to a significant extent by vertical fiscal imbalance. While States have the major service delivery responsibilities including in school education, they rely on substantial transfers from the Commonwealth, which raises the majority of the tax revenue. This resulted in a large number of specific purpose payments often with 'input controls'. One such example was the requirement for schools to have a flagpole carrying the Australia

* The author would like to acknowledge the support of Dr Sara Glover (Executive Director, Research and Analysis Division, Victorian Department of Education and Early Childhood Development), in the preparation of this chapter.

¹ Peter Dawkins was the former Secretary, Department of Education and Early Childhood Development, Victoria. He is currently Vice-Chancellor and President of Victoria University.

flag (Australian Government Programs for Schools, Quadrennial Administrative Guidelines 2005–2008).

Following the election of the Rudd Labor government in 2007, a process of reform was entered into, known as the National Productivity Agenda, following many of the ideas that had been proposed by Victorian Labor government in their proposed National Reform Agenda (DPC and DTF 2005). This involved the development of a national Education Agreement, involving payments from Commonwealth to State governments, linked with an outcomes framework; progress measures and targets; and an accountability and review framework, rather than detailed input controls. The National Education Agreement was supplemented with the National Partnership Agreements (on which, see Banks, Fenna and McDonald, this volume; O’Loughlin, this volume) to fund specific reforms and facilitate and/or reward States and Territories that deliver on these nationally significant reforms.

Thus benchmarking educational outcomes jurisdiction by jurisdiction became a key feature of the National Productivity Agenda, with a view to identifying those jurisdictions that implemented successful reforms that improved outcomes. This idea was along the lines of the former National Competition Policy. There is very little doubt that this agenda has motivated significant efforts in education systems to improve educational outcomes, and, in the authors’ view, has conceptual underpinnings as way of promoting educational progress in State systems in a world of vertical fiscal imbalance. However, it does bring with it a number of challenges that the Commonwealth and State governments have needed to confront in seeking to implement the policy successfully. This chapter discusses some of those challenges.

12.2 Background

There has been a strong history of benchmarking practices in school education across Australia. In 1993, the Heads of Government — now the Council of Australian Governments (COAG) — commissioned the *Report on Government Services* to help improve the effectiveness and efficiency of government services (see Banks and McDonald, this volume). COAG confirmed in late 2009 that the *Report on Government Services* should continue to be the key tool to measure and report on the productive efficiency and cost effectiveness of government services.

This framework of performance indicators aims to provide comparative information on the equity, efficiency and effectiveness of Commonwealth and State and Territory government services. The performance information promotes transparency and accountability; identifies areas of strong or poor performance; promotes

learning across governments; and creates an incentive to improve the performance of government services.

More recently, in 2008 the COAG Reform Council (CRC) was established to assist COAG drive its reform agenda by strengthening public accountability of the performance of governments through independent and evidence-based monitoring, assessment and reporting (see O’Loughlin, this volume).

The CRC reports on:

- the performance of the Commonwealth, States and Territories in achieving the outcomes and performance benchmarks specified in the National Agreements
- whether predetermined performance benchmarks have been achieved under National Partnerships.

In the case of National Partnerships, the CRC is the independent assessor of whether predetermined milestones and targets have been achieved. The assessment of each jurisdiction’s performance is reported publicly and a decision to make reward payments is based on this independent CRC assessment.

The latest development in benchmarking in School Education in Australia saw all Education Ministers agree to the publication of school information on the *My School* website. *My School* was launched in January 2010 by the Australian Curriculum and Assessment Authority (ACARA). ACARA was established by Commonwealth legislation in 2009, and is an independent authority responsible for the development of a national curriculum; a national assessment program; and a national data collection and reporting program that supports 21st century learning for all Australian students. ACARA has responsibility for publishing nationally comparable data on all — almost 10 000 — Australian schools on its *My School* website. Data on each school’s performance and factors relating to performance are provided. These include national testing in literacy and numeracy results; school attainment rates; student background characteristics; and information about each school’s teaching staff and income. Schools can compare their results in national literacy and numeracy tests with the results of other schools that serve similar students. They can also compare their progress against that of others schools that had the same starting point in 2008.

All results can be compared with results in statistically similar schools (that is, schools with similar student populations) across the nation. The *My School* website has been developed so that there is greater transparency and accountability about the performance of schools, and parents and the community have access to this information about their child’s school. The greatest potential value of this form of benchmarking lies in the support it provides for productive discussions between

schools that are doing better in similar circumstances to help them review and improve their own practices.

12.3 National Productivity Agenda and benchmarking school education

COAG agreements in late 2008 — the *Intergovernmental Agreement on Federal Financial Relations* and the *National Education Agreement* — resulted in all governments agreeing to a common framework for reform in education. An outcomes framework was developed and agreed to, establishing a set of aspirations, outcomes, progress measures and future policy directions to guide education reform in Australia — including a focus on improving outcomes for indigenous children and young people as well as students from low socio-economic backgrounds.

The underpinning idea of the National Productivity Agenda is that investment in human capital in the form of education and training raises work force participation and productivity, and therefore economic growth, and in turn government tax revenue (PC 2006; Dawkins 2010). This creates a virtuous circle.

In relation to school education, national and international evidence reviewed by the Productivity Agenda Working Group pointed to literacy and numeracy as a key determinant of school retention and subsequent workforce success, school completion as a major determinant of labour force participation, and teacher quality as the main in-school determinant of student success.

Using international benchmarking (especially using PISA data) it was also found that while the average performance of Australian students (for example, in literacy and numeracy at aged 15) is quite high by international standards, the performance of students from low socio-economic backgrounds was mediocre by international standards. Thus, key policy thrusts in the National Partnership Agreements included a focus on raising teaching quality, on raising literacy and numeracy and on improving outcome in schools with disproportionate numbers of students from low socio-economic backgrounds.

The very large gap in outcomes between indigenous and non-indigenous students was also a major focus. Consequently, the benchmarking that was agreed upon, was not only for average outcomes especially in literacy and numeracy and in school (year 12) completions, but also closing the gap between indigenous students and all students and between low-socio-economic background students.

12.4 What to benchmark

Determining *what* to benchmark is largely determined by the *purposes* of benchmarking. Benchmarking can be about promoting accountability and being transparent to the public about results. It can also be a practice that systematically evaluates relative strengths and weaknesses, searches for the evidence of improvements elsewhere, and more importantly how this was achieved (see Fenna, this volume). Benchmarking in the schools system aims to be both an assessment device and a learning tool.

Benchmarking in school education provides a mechanism of comparing the performance of the States and Territories. Currently these comparisons take different forms. These include the comparison of educational outcomes for children and young people; the comparison of inputs (for example, expenditure and teacher-student ratios); and the comparison of reforms and initiatives that have yielded improved results. Each of these different forms of benchmarking has their merits and value.

Benchmarking outcomes

The primary focus of the National Productivity Agenda is to benchmark outcomes because of the evidence that improving educational outcomes improves participation and productivity. Benchmarking outcomes compares performance between jurisdictions or cohorts on the actual results achieved by children and young people in each jurisdiction.

An example of this is benchmarking student reading results on the National Assessment Program — Literacy and Numeracy (NAPLAN) (figure 12.1).

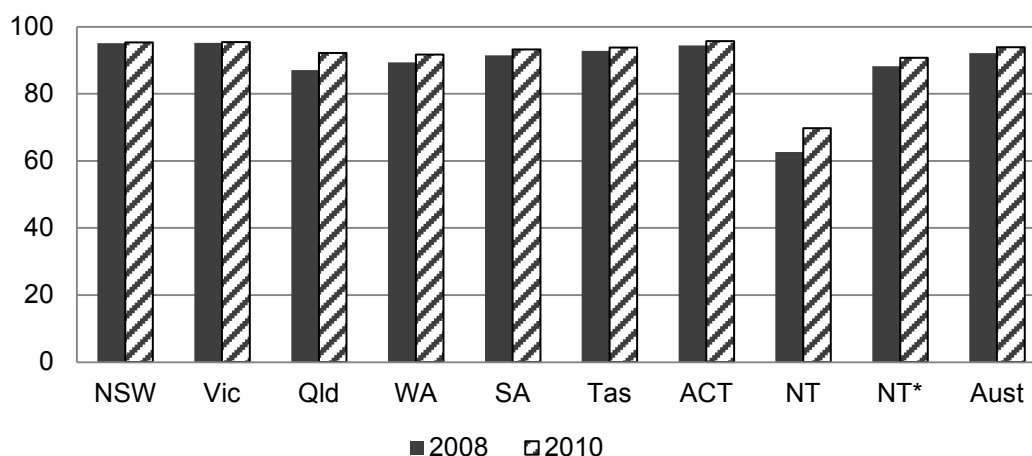
At face value, there is little difference between jurisdictions in these Year 3 reading results apart from the results for Indigenous students in the Northern Territory. Furthermore, with less than 10 per cent of children in Year 3 below the National Minimum Standard in reading one might question the value of this level of comparison. Other comparisons can also be made such as the proportion of students in the top or lower levels of performance and improvements of these over time.

Similarly comparison of the performance of different cohorts of students can be made. These include indigenous students; male and female students; students living in remote, very remote and metropolitan regions; students with language backgrounds other than English; and socio-economic status.

Most importantly we must be clear about the policy objectives and the outcomes we are trying to achieve and benchmark accordingly. Benchmarking whole year levels will mask differences and gaps in performance between cohorts of students and the

distribution of outcomes at the student and school level. It is possible, for instance, that improvements in student outcomes may occur at a jurisdictional level, yet not be apparent among low SES students.

Figure 12.1 **Year 3 reading: proportion of students at or above the national minimum standard**



* 'NT' includes all Australians resident in the NT. 'NT*' includes only non-Indigenous Australians resident in the NT.

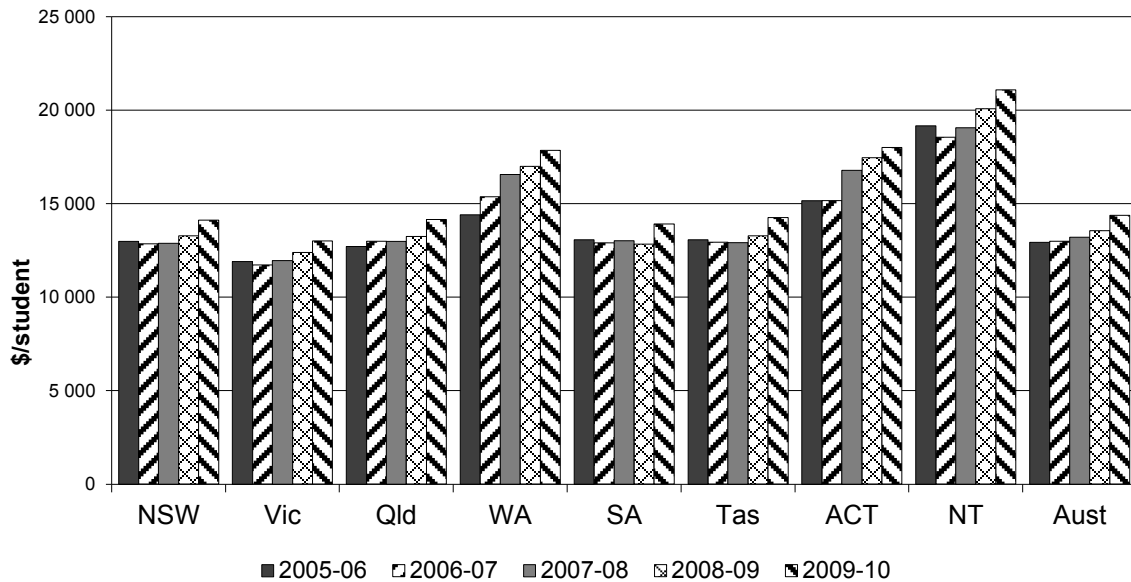
Source: COAG Reform Council (2011).

However, once we start benchmarking for sub-populations of students, issues of measurement error emerge. If this type of outcomes-benchmarking is used for evidence-informed decision-making then it is more likely that margins of error can be tolerated. Once external publications, performance judgments, and rewards or sanctions are applied to these types of benchmarking activities, there is potential to undermine or limit the ambition of the reform and/or the target group to be measured (Fenna, this volume). Attention and debate turn to issues of data and measurement, rather than policy and strategy. We return to this important issue in benchmarking later in this chapter.

Benchmarking inputs–outputs

Perhaps the most longstanding benchmarking practice in school education has been on benchmarking inputs and outputs. Examples include the recurrent expenditure per full-time equivalent student and student–teacher ratios.

Figure 12.2 **Government real recurrent expenditure per full time equivalent student, government schools (2009-10 dollars)**



Source: Steering Committee for the Review of Government Service Provision (2012).

The variation on government expenditure between jurisdictions is evident. What conclusions can be drawn from these benchmarked data? Does higher expenditure mean a less efficient system, or might it mean simply that the cost of educating students in more remote areas of Australia is higher and vice-versa does lower expenditure mean a more efficient system and greater economies of scale can be achieved in more densely populated areas? There is the potential for quite misleading conclusions to be drawn.

Similar issues arise with Student–Teacher ratios and average class size data. There is evidence that government investment in reducing class size is less effective than investments in improving teacher quality (Jensen 2010).

Such benchmarking creates a public pressure to maintain low class sizes without necessarily improving the quality of teaching and the outcomes for students. Furthermore, in jurisdictions with devolved decision-making regarding hiring of staff, schools may decide to employ a broader range of non-teaching staff to support individual students and families and provide necessary support for teachers. Benchmarked data on teacher numbers alone will not take these variations into account.

Benchmarking reforms

A question for policy makers is to understand how school systems improve performance. This type of benchmarking analyses the reform elements or sets of interventions of different systems that have led to significant gains in student outcomes as measured by national and international assessments. McKinsey & Co. (2007), for example, have undertaken international comparisons of school systems and subsequently Barber (2009) developed an empirically based framework for assessing the progress of different systems on a number of key dimensions.

Table 12.1 **Benchmarking system reform: nine characteristics**

<i>Standards and Accountability</i>	<i>Human Capital</i>	<i>Structure and Organisation</i>
<ul style="list-style-type: none"> • Globally-benchmarked standards • Good, transparent data • Every child on the agenda always in order to challenge inequality 	<ul style="list-style-type: none"> • Recruit great people and train them well • Continuous improvement of pedagogical skills and knowledge • Great leadership at school level 	<ul style="list-style-type: none"> • Effective, enabling central department and agencies • Capacity to manage change and engage communities at every level • Operational responsibility and budgets significantly devolved to school level

Source: Barber 2009.

Barber's framework incorporates three key themes which in his analysis have been the key to successful school systems. First there need to be rigorous performance standards against which schools and their students are to be assessed, and there needs to be appropriate levels of accountability for that performance. However, this cannot be successful without the second theme, which Barber calls the human capital in the system. That is, there needs to be a capacity building agenda to build an effective school workforce to be able to achieve the performance standards. Third, school systems need to be structured on a way that takes advantage of being a system, whole devolving appropriate responsibility to the school level. Under each of the three themes Barber identifies three characteristics that need to be present. This provides a basis for a school system to assess itself against what Barber concludes from his analysis is best practice for a school system.

Furthermore it is possible to benchmark the delivery of reforms or *how* a system implements the interventions. By doing this, systems can evaluate the relative strengths and weaknesses of the policy-implementation or delivery chain.

While these different benchmarking practices in school education in Australia offer important information for the public and for policy makers, the interplay between transparency, accountability and improvement in different contexts create a number of major challenges that may have both intended and unintended consequences.

12.5 Issues in benchmarking

The foregoing analysis has identified a very important role for benchmarking in seeking to improve educational outcomes for school students. However, there are a wide range of challenges in undertaking successful benchmarking. Furthermore, when the stakes are raised and when benchmarking itself becomes an integral part of the incentive and reward system — as under the National Partnerships — these challenges can make it difficult for the policy to be implemented successfully.

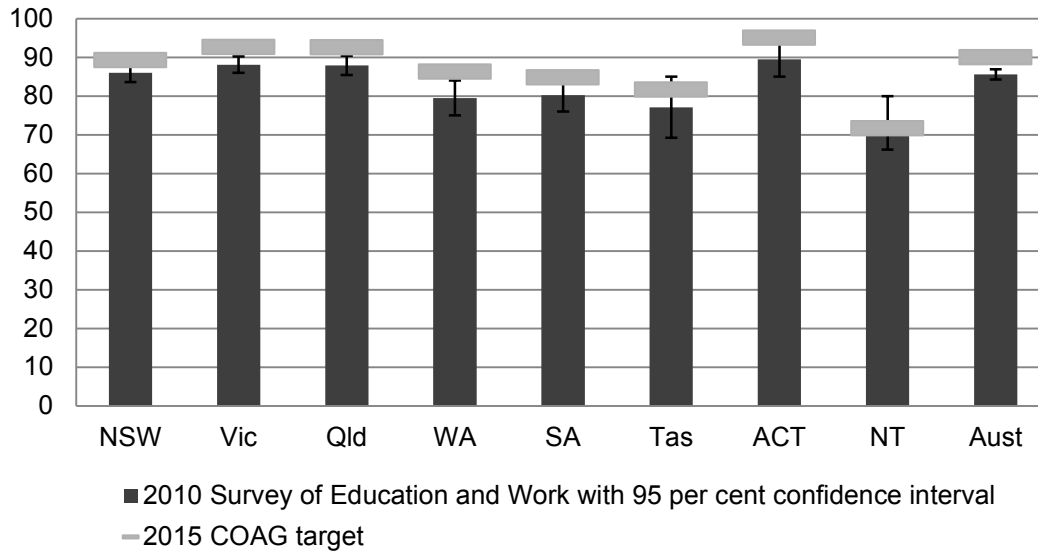
In this section we identify some of these challenges especially in relation the problems that have arisen under the National Education Agreement.

Measurement

Benchmarking for National Partnerships and My School requires nationally consistent and comparable data. The National Assessment Program — Literacy and Numeracy (NAPLAN), an annual national test for students in Years 3, 5, 7, and 9, is used for benchmarking literacy and numeracy. There are limitations of measurement in this assessment of students at the very top and bottom ends of performance. There is also measurement error. When it comes to ‘judging’ improved performance for sub-populations of students as well as comparing the performance of schools, the problem of measurement error cannot be discounted.

Similarly the only national comparable data for school attainment (year 12 or equivalent certification) is a national survey of education and work conducted by the Australian Bureau of Statistics. This is an annual sample survey and useful at a national level. However, when results are disaggregated by jurisdiction, large confidence intervals emerge. These make it difficult to compare jurisdictions; impossible to assess whether results in jurisdictions are improving over time; and relatively meaningless in trying to understand what is happening to Indigenous students, students from low socio-economic backgrounds and students in regional and rural locations in different States and Territories.

Figure 12.3 20–24 year olds with year 12 or equivalent or certificate II, per cent



Source: ABS (2011), Survey of Education and Work 2010.

Context

When benchmarking educational outcomes, it is important to take into account the different contexts. Thus, students from lower socio-economic backgrounds, for example, tend to underperform relative to students from higher socio-economic backgrounds. On the one hand, these effects need to be taken into account when benchmarking, and rewarding schools and school systems. On the other hand, one of the aims of school improvement is to reduce those effects, so benchmarking should not assume that these affects are set in stone.

Levels versus changes

When benchmarking measures of, for example, literacy and numeracy, it is important — partly because of the different contexts mentioned above — to look both at levels of student performance and changes in them. Thus it would have been unreasonable under the NAPLAN testing regime to expect that students in the Northern Territory will perform as well as the students from the ACT. Thus, in providing incentives and rewards the focus has been on making improvements in the levels. However, this raises further questions. For example, is it reasonable to expect students from the Northern Territory to improve faster than students from the ACT or vice versa?

When examining the trajectory of schools on the *MySchool Website*, if a school is already a high achiever, especially taking into account their socio-economic mix, how serious is it if in a particular year they appear to deteriorate? Should schools with very poor performance, or systems containing those schools, that improve marginally, be rewarded for so doing?

In putting forward targets for rewards, if a system is showing declining literacy or numeracy, should they be rewarded simply for slowing the decline, or should they actually achieve a real improvement?

These are very difficult questions to which there is no simple answer.

What outcomes to focus on

The main foci of the National Education Agreement have been literacy, numeracy and year 12 attainment. All these outcomes can be strongly defended as important and it is very good when outcomes can be improved in these areas.

Having said that, there is a very legitimate argument that there are other objectives of school education. Indeed the *Melbourne Declaration on Educational Goals for Young Australians* identified an array of outcomes that we should seek to focus on under the general headings of ‘Successful Learners’; ‘Confident and Creative Individuals’; and ‘Active and Informed Citizens’. Indeed there is a growing literature on 21st century skills that will be required by students graduating from our education systems (see for example Wagner). In general, improved literacy and numeracy are likely to be a significant input into many of these skills. On the other hand, if schools focus overwhelmingly on improving measurable outcomes in literacy and numeracy, and pay scant attention to other areas such as teamwork, problem solving, cross-cultural and communication skills, this may not be to the benefit of the students. As Fenna (this volume) notes, ‘teaching to the test’ (and ‘neglecting the broader suite of often less tangible or immediate desiderata’) is a perennial risk in performance monitoring regimes.

It is important to develop useful measures of these other attributes and use them in improving education and benchmarking schools and systems, but similar issues will arise about what weights to apply to the different measures.

Causality

When seeking to improve educational outcomes with the assistance of benchmarking, it is important to develop an understanding of what causes

improvements in outcomes. This requires sophisticated and in-depth analysis and evaluation. This can include multivariate statistical analysis, but also appropriate forms to share information about ideas and best practice.

When using targets to reward performance, the reward will flow whether the cause is anything to do with change in policy or practice. This may not be a problem provided it does create the right incentives and policy and practice does improve one way or another. However, it is important to undertake evaluation of the targets and their effects and refine and improve them over time.

Timeframes and orders of magnitude for expected improvement

To date, it is unclear what a reasonable timeframe for system improvement is. There has been little benchmarking on actual improvement taking into account different starting points. For a system to improve a mean score or lift a proportion of students above a particular standard takes time, and the evidence linking interventions to improvement over time remains limited.

Furthermore, improvements over time can be misleading. A feature of NAPLAN results is that improvements of systems, schools and students with lower starting points are likely to be greater than improvements observed from higher starting points. Unless some form of relative improvement is taken into account, absolute improvement scores can be quite misleading.

Hattie suggests that benchmarking in schooling should focus attention on the growth in student learning. For this to occur, education systems need to provide the tools and incentives to monitor individual student's progress, rather than relying on standardised scores and minimum standards.

Commonwealth–State relations and the use of benchmarking

The use of benchmarking to provide incentives and rewards to State and Territory jurisdictions for improved educational outcomes was a feature of the National Reform Agenda proposed by Victoria, which was ultimately adopted by COAG on the National Productivity Agenda. In Victoria's proposal the outcomes and measures would be developed and implemented in a collaborative federalist model, and administered by a federal entity reporting to COAG. In practice the Commonwealth Government who provided the funds for the incentives and rewards, itself runs the incentive and reward mechanisms albeit with some consultation and support from the COAG Reform Council, which provides periodic reports on the progress of systems in improving educational outcomes. This

approach brings with it the risk that it can become a top-down coercive performance monitoring model, rather than a collegial model which is more genuinely cooperative in nature with mutual goal development and standard setting, which would perhaps be more focussed on learning and improvement.

12.6 Conclusion

This chapter has canvassed a wide range of issues relating to benchmarking in school education. In a world of increasing transparency and accountability, good measurement and successful benchmarking, is a critical aspect of good educational policy. In Australia, through the National Education Agreement, and related developments, such as national testing of literacy and numeracy, benchmarking has become firmly embedded in national policy through the setting of incentives and rewards for state and territory jurisdictions.

This represents a very significant development in evidence based policy in Australia, and in the authors' view, continues to have great potential as a way of managing effective Commonwealth–State relations in a world of shared responsibility for education and vertical fiscal imbalance.

Nonetheless it brings with it significant challenges with which Australian policy makers are grappling. The problems are unlikely ever to be 'solved' in a permanent way — like most policy problems. But over time it is to be expected that we will learn from experience and make significant progress in the bid to promote educational outcomes in Australia. It is very important that the use of data and benchmarking to create incentives and rewards, is only one part of a broader approach to school improvement, that keeps schools, school systems and governments, focused on the essence of improvement, rather than just on measurement issues alone.

References

- Barber, M. 2009, *School Leadership and Whole System Reform*, presentation to EU Education Ministers Event, Sweden, 24 September 2009.
- COAG Reform Council 2011, *Education 2010: Comparing performance across Australia*, Statistical supplement, Report to the Council of Australian Governments, Sydney, September.
- Dawkins, P. 2010, 'Institutionalising an evidenced-based approach to policy making: the case of the human capital reform agenda', Chapter 11 in *Strengthening Evidence-Based Policy in the Federation*, Roundtable

Proceedings, Productivity Commission, Canberra, 17-18 August 2009, Volume 1.

DPC and DTF (Department of Premier and Cabinet and Department of Treasury and Finance) 2005, *Governments Working Together: A Third Wave of National Reform – A New National Reform Initiative for COAG: The Proposals of the Victorian Premier*, DPC, Melbourne, Victoria.

Hattie, J. 2009, *Visible Learning: a synthesis of over 800 meta-analyses relating to achievement*, Routledge, London.

Jensen, B 2010, *Investing in Our Teachers, Investing in Our Economy*, Grattan Institute, Melbourne.

McKinsey & Co. 2007, *How the World's Best Performing School Systems Come out on Top*, http://www.mckinsey.com/clientservice/socialsector/respurces/pdf/Worlds_School_Systems_Final.pdf.

PC (Productivity Commission) 2006, *Potential Benefits of the National Reform Agenda*, Report to the Council of Australian Governments, Research Paper, Productivity Commission, Canberra.

Steering Committee for the Review of Government Service Provision 2012, *Report on Government Services 2012*, Productivity Commission, Canberra.

Wagner, T., 2008 *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills Our Children Need – And What We Can Do About It*, Basic Books, New York.