NSW GOVERNMENT Education & Communities

Office of Education

LITERATURE REVIEW

School improvement frameworks: The evidence base

Centre for Education Statistics and Evaluation

School improvement: background

Continuous, systemic school improvement is increasingly seen as essential by education systems. The focus on school improvement is driven in part by a growing awareness of international educational performance (through, for instance, the Programme for International Student Assessment, or PISA), and major comparative works published by organisations such as the OECD and McKinsey & Company. One recent OECD report argues that assessment and evaluation of students, teachers and schools are 'becoming critical' to establishing high performance and providing feedback, with the ultimate goal of improving student outcomes[1].

In his foreword to the 2007 McKinsey & Company paper, *How the world's best-performing school systems come out on top*, Andreas Schleicher explained the economic imperative sitting behind the drive for improvement:

> *The capacity of countries … to compete in the global knowledge economy increasingly depends on whether they can meet a fast-growing demand for high-level skills. This, in turns, hinges on significant improvements in the quality of schooling outcomes and a more equitable distribution in learning opportunities*[2].

Although it is widely accepted as necessary, systemic and continuous improvement is also acknowledged to be a complex process, requiring action over many domains. Many large initiatives for school improvement fail because they do not change day-to-day school practices, which are 'recognised as remarkably impervious to, and self-protective against, fluctuating external policies and agendas'[3].

A related challenge is sustaining change once it has been enacted: sustainability is essential if Australia wants to become and remain a top educational performer. Sustaining change means building capacity within schools, to ensure that teachers and schools are adaptive, capable of continuous learning, and can take charge of change[4]. As part of this process, schools and districts must work together to share best practice[5]. This 'systemic' approach is not new, however Barber and Fullan have argued that what is needed is a shift from systems *thinking* to systems *action* — for the strategic, powerful pursuit of improvement in practice[6].

Many large initiatives for school improvement fail because they do not change day-to-day school practices.

There is evidence that substantial, long-lasting change is possible. McKinsey & Company's 2010 report, *How the world's most improved school systems keep getting better*, identifies a range of improved education systems from across the world that have made sustained improvements from a wide range of starting points[7]. For instance, Singapore has transformed its education system from 'fair' to 'great' in a twelve-year period; and Ontario, Canada moved from 'good' to 'great' within the space of ten years[8].

The McKinsey report also provides guidance as to the kinds of practices improving educational systems have undertaken, to foster that improvement. These practices include building teacher and principal technical skills; student assessment; use of data systems; facilitation of improvement through policy and law as well as revision of curriculum and standards; and ensuring appropriate reward and remunerations structures for teachers and principals[9].

In recent years, many school systems have developed school improvement or performance frameworks as a means of:

- identifying the core components of the school improvement process
- supporting schools through this process, and
- assessing schools' performance against core components.

1. *OECD 2013, Synergies for better learning: An international perspective on evaluation and assessment,* p.13.
2. McKinsey & Company 2007, *How the world's best-performing school systems come out on top*, report prepared by M Barber and M Mourshed, p.8.
3. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2012a, *Measuring and rewarding school improvement*, paper prepared by G Masters, p.1.
4. L Stoll 2009, 'Capacity building for school improvement or creating capacity for learning? A changing landscape', *Journal of Educational Change*, vol.10, p.117.
5. NSW Department of Education and Training 2005, *Building a more responsive system of public education*, companion paper 5, prepared by M Fullan.
6. M Barber and M Fullan 2005, 'Tri-level development: It's the system', *Education Week*, March 2.
7. 'Sustained improvers' are defined as education systems that have seen five years or more of consistent improvements in student performance, across multiple data sets and subjects. McKinsey & Company 2010a, *How the world's most improved school systems keep getting better*, report prepared by M Mourshed, C Chijioke and M Barber, p.11.
8. Major educational reforms began in 2003 in Ontario. McKinsey & Company 2010a, p.19.
9. McKinsey & Company 2010a, p.20.

All states and territories in Australia have developed such frameworks. Australian interest has sharpened recently, with the development of a National School Improvement Tool (NSIT), intended for implementation as part of the National Plan for School Improvement.

There is evidence that substantial, long-lasting change is possible.

Many of these frameworks use standards as policy levers. These are useful as they provide a clear articulation of good practice; they support self-reflection and assessment; and, as they are not relative, they support system-wide improvement. For instance, the NSIT describes school performance at 'Low', 'Medium', 'High' or 'Outstanding' levels. This allows schools to 'make judgements about where they are on their improvement journeys, to set goals and design strategies for improvement, and to monitor and demonstrate school improvement over time'[10].

This paper considers the research literature related to school improvement frameworks, to identify the core components and processes of such frameworks, and to assess evidence of their efficacy. While there is a substantial body of research bearing on the development of improvement frameworks, the evidence regarding the effect of frameworks on student outcomes is comparatively slight and inconclusive. This is because it is much easier to describe a framework than it is to test whether its implementation improves student outcomes. However, as outlined below, many of the individual components of school improvement frameworks are underpinned by robust evidence.

10. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2012b, *National School Improvement Tool*, prepared by the Australian Council for Educational Research, p.1.

How is improvement measured?

The international evidence indicates that school improvement is best measured with reference to both student outcomes and school practices or processes, rather than by focusing exclusively on one or the other. Raudenbush argues that no matter how sophisticated analysis of student outcome data is, there will always be limitations as to how much the data can tell us, claiming that 'to be successful, accountability must be informed by other sources of information … in particular, information on organizational and instructional practice'[11]. Similarly, Elmore argues that accountability systems must go beyond testing and regulation, to actively engage those who work in schools through 'explicit strategies for developing and deploying knowledge and skill in classrooms and schools'[12].

A dual approach (encompassing both outcomes and practices) is already evident in some school systems. For instance, the Nova Scotia School Accreditation Program (NSSAP) requires schools to select one area of practice alongside one area of student achievement for improvement. Working in 'professional learning communities', staff members collectively establish goals for these focal areas, while the school more broadly establishes strategies to meet them[13].

School practices

Although improvement frameworks vary in their structure and terminology, there is a core set of focal areas which consistently appear, notably:

- teaching[14]
- learning[15]
- leadership[16]
- a focus on school improvement[17]
- data analysis[18]
- community partnerships[19]
- student wellbeing[20].

School improvement is best measured with reference to both student outcomes and school practices or processes.

Often, these areas of interest are articulated in ways that emphasise particular types of effective practice. For example, the NSIT distils the research base on school improvement[21], to identify nine areas:

- explicit improvement agenda
- analysis and discussion of data
- a culture that promotes learning
- targeted use of school resources
- an expert teaching team
- systematic curriculum delivery
- differentiated teaching and learning
- effective pedagogical practices
- school-community partnerships.

The focus of this paper is learning, teaching and leadership, because these are common to most frameworks that have been developed worldwide.

11. S Raudenbush 2004, 'Schooling, statistics and poverty: Can we measure school improvement?' presented at the *William H Angoff Memorial Lecture Series, Educational Testing Service*, Princeton NJ, 1 April, p.37.
12. R Elmore 2006, 'Leadership as the practice of improvement', presented at the *International Conference on Perspectives on Leadership for Systemic Improvement*, London, 6 July, p.3.
13. C Wood and M Meyer 2011, 'Impact of the Nova Scotia School Accreditation Program on teaching and student learning: An initial study', *Canadian Journal of Educational Administration and Policy*, Issue 124, p.2.
14. Australia (National School Improvement Tool); Australian Capital Territory (School improvement framework); Queensland (Teaching and learning school improvement framework); Western Australia (The school improvement and accountability framework); Northern Territory (School accountability and performance improvement framework); Victoria (Accountability and improvement framework); United Kingdom (Ofsted); United States (New England Association of Schools and Colleges; AdvancED).
15. Australia; Australian Capital Territory; Queensland; Western Australia; South Australia (Improvement and accountability framework); Northern Territory; Victoria (Accountability and improvement framework; Performance and development culture); United Kingdom; United States.
16. Australia; Australian Capital Territory; Queensland; Western Australia; Northern Territory; Victoria (Accountability and improvement framework); United Kingdom; United States.
17. Australia; Queensland; United States (AdvancED).
18. Australia; Queensland; South Australia.
19. Australia; Queensland; Australian Capital Territory; Western Australia; Northern Territory.
20. Australian Capital Territory; Northern Territory; Victoria (Accountability and improvement framework); United Kingdom.
21. Commonwealth DEEWR 2012a, p.1.

Quality practices: The evidence base

The NSIT and other similar frameworks generally do not include reference to an explicit evidence base for each of their component elements. This is because such frameworks are typically trying to capture all the key drivers of school improvement, and to assess the evidence base for effective practices across all domains would amount to reviewing the entire evidence base for quality education provision. It also reflects the fact that strategies to improve learning, teaching or leading rarely occur in isolation, and thus their unique effects on student outcomes can be difficult to tease apart from other activities occurring at the school and system levels.

What the following section of this paper does is outline some of the practices across the domains of learning, teaching and leading for which there is the strongest quantitative evidence.

Effect sizes provide one useful measure of efficacy. They provide researchers with a scale capable of comparing outcomes from various diverse studies: a 'common expression of the magnitude of study outcomes for many types of outcome variables'. Hattie classifies programs with an effect size greater than 0.40 as having the greatest impact on student outcomes[22].

Quite often the impact of practices on student outcomes is not measured (or cannot be measured) in a manner rigorous enough to allow programs and practices to be compared on this scale. In these cases, other information may be available that suggests (but cannot prove) the efficacy of practices — for instance the fact that certain practices are undertaken in high-performing or improving countries.

Learning

It is widely accepted that the ultimate purpose of school improvement is to improve students' outcomes. Barber and Fullan refer to the consistent improvement of student outcomes, and the narrowing of achievement gaps between students, as the central 'moral purpose' of schools[23].

A culture of high expectations

Comparative research has drawn links between high expectations and high performance. Tucker notes that in Japan, academic achievement is perceived to be the result of hard work, not innate ability. Japanese teachers demand that students work hard and have high expectations of all students[24].

Empirical studies dating back to the 1960s have pointed to the impact of teacher expectations on student performance. For instance, in the 1965 'Pygmalion in the classroom' study, researchers Rosenthal and Jacobson told teachers that a group of randomly selected elementary school students could be expected to be 'growth spurters'. The teachers were told that the children had been identified through a new test. School-wide, the 'spurters' gained almost four IQ points more than the control group – a statistically significant difference. Amongst the younger students (Grades One and Two), the effects of teachers' expectations were dramatic: Grade One students gained over 15 IQ points more, and Grade Two students over nine points more, than their respective control groups (both findings were statistically significant)[25].

A recent report on young Australians' aspirations found that even when controlling for student background and prior achievement, students who intended to complete Year 12 were 20 to 25 per cent more likely to do; and students who planned on attending university were between 15 and 20 per cent more likely to do so[26].

22. Hattie's research (which measures effect size in terms of the standard deviation of educational performance), and educational research generally, tends to use Cohen's d to determine effect size. J Hattie 2009, *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*, Routledge, Oxon, p.19.
23. Barber and Fullan 2005, pp.34-35.
24. M Tucker 2012, *Surpassing Shanghai*, Harvard Education Press, Cambridge MA, p.103.
25. R Rosenthal and L Jacobson 1968, 'Pygmalion in the classroom', *The Urban Review*, vol.3, no.1, p.17.
26. J Homel and C Ryan 2014, *Educational outcomes: The impact of aspirations and the role of student background characteristics*, Longitudinal Surveys of Australian Youth: Research Report No. 65, National Centre for Vocational Education Research, p.27.

Performance difference between students of teachers with high and low expectations may be linked to discrete teacher practices. A very small-scale study conducted by Rubie-Davies (which did not examine student outcomes) found that teachers with high expectations more regularly oriented students to their lessons; used students' prior knowledge; provided students with instructions and explanations; modified lessons according to students' needs; provided feedback to students; and managed their students' behaviour in a positive manner[27].

Curriculum

It is important that high expectations are supported by schools in practice through the provision of a broad curriculum, including challenging subjects such as extension units. Australian students in rural Australia have reported (in a series of focus group interviews) that reduced subject choice can create a barrier to forming career aspirations[28].

American research has found that placement in lower-ability, standard or advanced classes impacts on the academic achievement of students who start with similar pre-placement achievement. Oakes found that students lost an average of 2 Normal Curve Equivalents (NCEs) when placed in lower-ability classes[29]; while those placed in standard and advanced classes gained on average 3.5 and 9.6 NCEs, respectively[30]. Oakes argued that teachers expected less of students in the lower-track classes, who were disproportionately African American or Latino. Students in these classes were exposed to lower levels of curriculum and instruction and provided with less access to the kinds of courses that would qualify them for college entrance[31].

Responsive teaching practices

Responsive practices begin with feedback, which Hattie identifies as being 'among the most powerful influences on achievement'[32]. A review of meta-analyses conducted by Hattie and Timperley found that the average effect size of feedback was 0.79, an effect size comparable to that of students' prior cognitive ability (0.71) and direct instruction (0.93)[33].

Hattie emphasises that feedback is not only something provided by teachers to students; feedback is more powerful when teachers are open to feedback from their students[34].

Researchers have pointed to the advantages of curriculum differentiation, which involves building upon individual students' prior knowledge through 'tiered instruction or alternative curriculum'[35]. Tieso found that curriculum differentiation, in combination with flexible small groupings within classes, had medium to high effect sizes for all students, with the greatest effect for higher-performing students (0.83, compared to 0.42 for students who had scored in the middle of the performance range, and 0.29 for the treatment group that had scored lowest[36]).

Accelerated instruction is one form of curriculum differentiation used to meet the needs of high-achieving students. It includes skipping a year or compressing a curriculum (for example, completing four years of study within three years[37]). In their meta-analysis, Kulik and Kulik found that the average effect size of accelerated instruction was 0.88 compared to students of the same age who were not accelerated[38].

27. C Rubie-Davies 2007, 'Classroom interactions: Exploring the practices of high- and low-expectation teachers', *British Journal of Educational Psychology*, vol.77, pp.289–306.
28. N Alloway and L Dalley-Trim 2009, '"High and dry" in rural Australia: Obstacles to student aspirations and expectations', *Rural Society,* vol.19, no.1, pp.56-57.
29. Normal Curve Equivalents are normalised, standardised scores with a mean of 50, developed to measure gains over time. C Mertler 2002, *Using standardized test data to guide instruction and intervention,* ERIC Digest, ERIC Clearinghouse on Assessment and Evaluation, p.3. NCEs can be converted to standard deviations by dividing the NCE by 21.06.
30. J Oakes 1995, 'Two cities' tracking and within-school segregation', *Teachers College Record*, vol.96, no.4, p.689.
31. Oakes 1995, p.687.
32. Hattie 2009, p.173.
33. J Hattie and H Timperley 2007, 'The power of feedback', *Review of Educational Research*, vol.77, no.1, p.83.
34. Hattie 2009, p.173.
35. C Tieso 2005, 'Effects of grouping practices and curricular adjustments on achievement', *Journal for the Education of the Gifted*, vol.29, no.1, p.64.
36. Tieso 2005, p.76.
37. J Kulik and C Kulik 1984, 'Effect of accelerated instruction on students', *Review of Educational Research*, vol.54, no.3, p.412.
38. Kulik and Kulik 1984, p.415.

Teaching

Teachers have the largest in-school influence on student outcomes[39]. This section of the paper identifies some aspects of teaching that research suggests have substantial impacts on student learning.

Collaborative practice

Collaborative practice is reported by educational experts to be one of the most important factors in improving schools, and to operate as a form of peer accountability[40]. Stoll identifies the role of collaboration in improvement, noting that: 'different parts of the system need to be aligned to provide a coherent and consistent picture and strategy for improvement, and people with diverse roles in the system will have to connect and learn together'[41].

Consistently high-performing countries build capacity through collaboration. In high-performing Asian countries, teachers are viewed as researchers. They observe other teachers' classes and provide feedback[42]. Tucker describes practices such as peer observation, mentoring and demonstration lessons as so common in China they are 'taken for granted', and argues that they play an important role both in quality assurance, and in professional development[43].

In Australia, the Melbourne Graduate School of Education has introduced a Masters of Teaching that is heavily influenced by the idea that teaching should be 'recognised as an academically taught, clinical practice profession'. One component of such professions is that they form a 'professional community that monitors quality, distributes knowledge and creates standards of practice'[44]. A similar model, Quality Teaching Rounds[45], is being evaluated by the University of Newcastle[46].

While there is broad support for and interest in collaborative practices such as networks and professional communities, quantitative evidence linking such practices to student outcomes is limited.

Louis and Marks published a study in 1998 which sought to assess the strength of the association between professional community and student academic outcomes. Professional community was defined as including shared values, focus on student learning, collaboration, 'deprivatised' practice and reflective dialogue[47]. The authors reported an effect size of 0.26 between professional community and student outcomes, to a high level of statistical significance[48], but the nature of the study meant that it could not show that the professional community had 'caused' these improvements[49].

An earlier study by Stevens and Slavin found that a cooperative elementary school model which increased teacher and principal collaboration resulted in large academic effect sizes (ranging from 0.25 to 0.85) for special education students in the second year of the study[50], compared to students in control schools. However, as a number of other measures were implemented as part of the model it is difficult to ascertain the precise impact of the model's component parts.

Professional learning and development

Teachers in Australia report almost universal access to professional development opportunities. In the 2013 OECD Teaching and Learning International Survey (TALIS), 97 per cent of Australian teachers reported undertaking some professional development activities in the past 12 months, compared to an average of 88 per cent in the 34 participating countries. However, Australian teachers appear to spend less time on professional development than teachers in other countries: Australian teachers reported spending an average of four days in the past 12 months attending courses and workshops, compared to eight days in other countries. Australian teachers also report spending far less time on other common development activities, such as conferences, visits, or observation visits to other schools[51].

39. See, for instance, McKinsey & Company 2007, p.12; OECD 2009, *Evaluating and rewarding the quality of teachers: International practices*, p.13.
40. McKinsey & Company 2010a, p.75.
41. Stoll 2009, p.124.
42. B Jensen 2012, *Catching up: Learning from the best school systems in East Asia*, Grattan Institute Report No. 2012-3, p.23.
43. Tucker 2012, p.29.
44. J Alter and J Coggshall 2009, T*eaching as a clinical practice profession: Implications for teacher preparation and state policy,* National Comprehensive Center for Teacher Quality, New York; cited in L Davies et al, 2013, 'Masterly preparation: embedding clinical practice in a graduate pre-service teacher education programme', *Journal of Education for Teaching: International research and pedagogy,* vol. 39, no. 1, pp.94-95.
45. J Gore, N Mockler, M Smith and J Bowe 2012, *Response to 'Great teaching, inspired learning'*, submission in response to the NSW Department of Education and Communities discussion paper: *Great Teaching, Inspired Learning*, University of Newcastle, viewed 8 September 2014, www.schools.nsw.edu.au/media/downloads/news/greatteaching/submissions/university-of-newcastle.pdf.
46. The University of Newcastle, *Professor Max Smith, Grants and funding,* viewed 8 September 2014, www.newcastle.edu.au/profile/maxwell-smith#profile-grants-funding.
47. K Louis and H Marks 1998, 'Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools', *American Journal of Education*, vol.106, no.4, p.539.
48. Louis and Marks 1998, p.549.
49. Note: the actual effect size may differ from this reported effect size.
50. R Stevens and R Slavin 1995, 'The cooperative elementary school: Effects on students' achievements, attitudes and social relations', *American Educational Research Journal*, vol.32, no.2, p.338.
51. OECD 2013, *Australia: Key findings from the Teaching and Learning International Survey (TALIS)*, viewed 8 September 2014, www.oecd.org/australia/TALIS-2013-country-note-Australia.pdf.

Most evaluations of professional development programs tend to focus on the impact of these programs on teachers, rather than considering their impact on students. One small-scale Victorian study of teacher training (in behaviour management strategies) demonstrates the problem with this approach. It found that even when teachers reviewed professional development programs favourably, they did not necessarily apply the strategies they had learned in the classroom[52].

Some research is available demonstrating the impact of professional development on student outcomes. For instance, Scher and O'Reilly's meta-analysis of professional development for K-12 maths and science teachers found that maths-focused professional development emphasising content knowledge, how to teach that content, and how students learn generated larger effect sizes on student outcomes (0.56) than programs focusing on pedagogy alone (0.07)[53].

Data skills and use

One particularly effective area of professional development is data skills and use. Timperley published a paper in 2009 reporting that a professional development program for teachers that focused on the interpretation and use of assessment information resulted in student achievement gains accelerating at twice the expected rate. For all schools that focused on writing, the average effect size was 1.20; for reading, 0.92. Gains were found to be greatest for the lowest-performing 20 per cent of students: effect sizes were 2.25 in writing and 1.90 in reading for these students[54].

Teacher content knowledge

In some high-performing countries, strong content knowledge is either a pre- or co-requisite of teaching. In Finland, all teachers must obtain a master's degree prior to receiving their teaching qualification. The master's degree includes a research-based paper, on a topic either in the student teacher's subject area or on pedagogy[55].

In Singapore, initial teacher education has a strong focus on subject content, and graduates are experts in their subject area. For example, mathematics teachers graduate with the equivalent mathematical knowledge of students studying pure maths[56]. Teachers in Shanghai also form research groups, regularly publishing papers on areas they think will improve student learning in district or municipal level publications[57].

Some program evaluations indicate the importance of teachers' content knowledge. A 2009 analysis by Boyd and others found that content-specific coursework in mathematics was positively associated with teachers' value-added scores in their second year of teaching, and had a small, inconsistent impact in first year; findings were similar but smaller for English courses. The authors suggested that content knowledge may be more important in teachers' second years, when they are more comfortable with day-to-day teaching practices. The authors urge caution in interpreting their results however, noting that research analysing the relationship between teacher-preparation programs and their effect on student achievement is 'still in its infancy'[58].

In Hattie's synthesis of more than 800 meta-analyses he argues that there is little support for the claim that teacher knowledge alone impacts on student outcomes[59]:

> *Experts possess knowledge that is more integrated, in that they combine the introduction of new subject knowledge with students' prior knowledge; they can relate current lesson content to other subjects in the curriculum; and they make lessons uniquely their own by changing, combining and adding to the lessons according to their students' needs and their own teaching goals*[60].

52. R Giallo and L Hayes 2007, 'The paradox of teacher professional development programs for behaviour management: Comparing program satisfaction alongside changes in behaviour management practices', *Australian Journal of Educational & Developmental Psychology,* vol.7, pp.108-119.
53. L Scher and F O'Reilly 2009, 'Professional Development for K-12 Math and Science Teachers: What do we really know?', *Journal of Research on Educational Effectiveness*, vol.2, p.230.
54. H Timperley 2009, 'Using assessment data for improving teaching practice', Paper presented at the Australian Council for Educational Research Conference, 16-18 August, viewed 8 September 2014, http://research.acer.edu.au/cgi/viewcontent.cgi?article=1036&context=research_conference; see also, H Timperley and J Parr 2009, 'Chain of Influence from policy to practice in the New Zealand literacy strategy', *Research Papers in Education*, vol.24, no.2, pp.135-154.
55. OECD 2011b, *Strong performers and successful reformers in education: Lessons from PISA for the United States*, p.125.
56. Jensen 2012, p.54.
57. Jensen 2012, pp.92-93.
58. D Boyd et al, 2009, 'Teacher preparation and student achievement', *Educational Evaluation and Policy Analysis*, vol.31, no.4, pp.431, 434-5.
59. Hattie 2009, pp.113-14.
60. J Hattie 2012, *Visible learning for teachers: Maximising impact on learning*, Routledge, New York, p.25.

Leading

School leadership has grown in importance over the past decade[61], as part of the trend towards devolution of school management to schools.

One UK study found that 97 per cent of schools in England rated good or excellent overall by the independent inspectorate (Ofsted) had leadership teams that were also rated good or excellent; however only 8 per cent of schools with leadership teams rated satisfactory or below were rated good or excellent[62].

Research supports 'instructional leadership' as the most effective style of leadership[63]. Instructional leaders focus on improving teachers' teaching. They have a clear set of teaching objectives and high expectations of students[64]. On the other hand, leadership styles that focus on inspiring staff ('transformational leadership') are generally accepted to be less effective. A 2008 meta-analysis conducted by Robinson, Lloyd and Rowe found that instructional leadership had an effect size of 0.42 on student outcomes (including non-academic outcomes); while transformational leadership had an effect size of 0.11[65].

The authors also identified a number of effective leadership 'dimensions', finding that 'promoting and participating in teacher learning and development' (where leaders are 'leading learners' in the schools, and act as a source of instructional advice) is associated with the largest average effect size, of 0.84[66].

Student outcomes

The ultimate goal of school improvement is improved student outcomes and some measurement of student outcomes is central to most school improvement frameworks. Jurisdictional frameworks differ in how they incorporate student performance data, though there are three commonly employed methods for measuring trends in student academic outcomes:

- Status or 'absolute' measures (comparing the performance of one grade level over time, for instance Year 5 reading between 2008 and 2013).
- Gain or growth models (examining the change in a cohort's scores between two points in time, for instance the increase in test scores between Year 3 in 2008 and Year 5 in 2010).
- Value-added measures (analysing student growth over time, and asking whether a school has added value over and above the performance expected given the school and student characteristics).

There are benefits and limitations attached to each of these approaches, which are explored in turn below.

Status measures

A status measure is the simplest measure of student outcomes. It measures the performance of one grade level over time — for instance, results in Year 5 reading between 2008 and 2013. It involves different cohorts of children each year and provides a 'snapshot' of achievement rather than a measure of the progress each cohort makes[67].

61. McKinsey & Company 2007, p.29. See also: McKinsey & Company 2010b, *Capturing the leadership premium: How the world's top school systems are building leadership capacity for the future*, report prepared by M Barber, F Whelan and M Clark, p.5; B Pont, D Nusche, H Moorman 2008, *Improving school leadership*, Volume 1: Policy and Practice, OECD Publishing, Paris, p.9; A Schleicher (ed) 2012, *Preparing teachers and developing school leaders for the 21st century: Lessons from around the world*, OECD Publishing, Paris, p.13.
62. McKinsey & Company 2007, p.29.
63. Hattie 2009, p.83.
64. V Robinson, C Lloyd and K Rowe 2008, 'The impact of leadership on student outcomes: An analysis of the differential effects of leadership types', *Educational Administration Quarterly*, vol.44, no.5, p.638.
65. Robinson, Lloyd and Rowe 2008, p.655.
66. Robinson, Lloyd and Rowe 2008, p.663.
67. Council of Chief State School Officers 2005, *Policymakers' guide to growth models for school accountability: How do accountability models differ?*, prepared by P Goldschmidt et al, p.7.

While accessible and intuitive, status measures raise a number of issues. First, as they involve different cohorts, they can fluctuate with changes in the student cohort. Secondly, these measures do not automatically consider students' or schools' contexts, such as socio-economic status (SES) or prior achievement. A status measure, used in isolation from other measures, may imply that 'all student success is attributable to the current school in the current year' and that 'students do not bring any "human capital" inputs with them'[68].

Despite the limitations of status measures, Masters prefers them for school-performance purposes. He argues that in the long term, year-to-year variation is less of a problem, and that status measures are more reliable than long-term averages of other measures (described below)[69]. The US-based Council of Chief State School Officers notes that the impact of a school on student outcomes may be acknowledged by explicitly incorporating students' backgrounds into a status model, or by using a growth model (see below)[70].

Gain or growth measures

Student gain or growth is an alternative measure of student performance that focuses on the change in a student's or cohort's scores between points in time[71]. O'Malley and co-authors argue that growth models are important because they 'conceptually align well with one of the fundamental goals of education – student learning … [which] implies change over time' and 'provide richer information on student learning than a single score at one point in time because they connect scores from multiple assessments'[72].

Growth models use schools' or students' initial achievement as the baseline. They take account of aspects of a student's background (including SES) insofar as they are reflected in differences in starting point — so, at least theoretically, all students start off on equal footing. Focusing on growth gives all schools — even those over- or under-achieving by a significant amount on other measures — ambitious but achievable targets for further improvement. As Goldschmidt and others point out, an advantage of growth models is that they:

> *… assume that student performance, and by extension school performance, is not simply a matter of where the school is at any single point in time, and a school's ability to facilitate academic progress is a better indicator of its performance*[73].

Growth measures do not capture the effect of any years outside of their frame of reference. For instance, measuring the gain a child makes between Year 1 and Year 2 does not take into account possibly substantial gains that the child may have made in Kindergarten. Such a measure may not recognise the work done in the year prior[74]. At the student level, students with high starting points might have different learning trajectories compared to those with low starting points, and rates of learning as well as learning outcomes may be affected by SES.

A further challenge for these models is that measuring growth requires vertically equated tests that allow students to be compared across year groups, and education systems need to have the capacity to track individual students. While assessment in New South Wales has vertically equated tests from Year 3 to Year 9 in NAPLAN (National Assessment Program – Literacy and Numeracy), growth measures cannot be directly calculated for two test times that use non-equated assessments, such as growth from Kindergarten (Best Start) to Year 3 (NAPLAN).

Value-added measures

Value-added measures add a further dimension to models based on student growth. The value-added estimate is the amount of growth the school adds to student performance relative to that achieved by the average school, after accounting for characteristics of students and schools, such as SES and school academic selectivity. This means that a school can achieve growth for its students, but if that growth is less than

68. Council of Chief State School Officers 2005, p.15.
69. Commonwealth DEEWR 2012a, p.11.
70. Council of Chief State School Officers 2005, p.14.
71. K O'Malley et al, 2011, *Overview of student growth models*, White Paper, Pearson Education, p.3.
72. O'Malley et al, 2011, p.4.
73. Council of Chief State School Officers 2005, p.5.
74. Raudenbush 2004, p.14.

the average for similar schools with similar students, its value-added estimate can be negative[75].

The primary strength of value-added measures is their capacity to take into account differences in student starting points and other contextual variables such as SES that impact on student learning outcomes, recognising the different workloads of schools serving different communities. A paper published by the Victorian Department of Education argues that value-added estimates take student background into account more than any other measure, providing a more accurate indication of the school's influence[76].

Value-added measures have been criticised on account of their comparative opacity. Raudenbush argues that a value-added system is not transparent to policy makers and the public, because data-analysis requirements will be significant[77]. Similarly, Masters maintains that value-added measures are difficult for schools and parents to interpret[78], and that it may be difficult for schools to decide how to respond to such a measure.

Nonetheless, value-added measures are the fairest way to assess the relative contribution of each school to the learning progress of its students. They are used and well-accepted in many jurisdictions, including Hong Kong and the New South Wales Catholic sector. It is worth noting that their use in these settings is mostly confined to the school-level, and mostly not publicly reported. More information on value-added measures, including the jurisdictions in which they are used is contained in the Centre for Education Statistics and Evaluation's technical paper, *Value added models for NSW schools*[79].

Acknowledging socio-economic status

One of the major criticisms of status measures, noted above, is that they discriminate against schools with high proportions of students from low-SES backgrounds. The advantage of growth and value-added measures, on the other hand, is that they incorporate students' starting points (including SES), to differing degrees.

While the rationale for accounting for SES when measuring school performance is obvious, there are two possible disadvantages to its incorporation. First, the effect may be to go beyond acknowledging the impact of SES to effectively lowering expectations for low-SES schools[80]. Secondly, if more advantaged schools attract more able teachers, then:

> *demographic and SES variables become proxies for teacher and school quality … if disadvantaged students are systematically assigned to less effective schools and teachers, inclusion of SES as a control can mask genuine differences in school and teacher quality*[81].

Thirdly, it could be argued that funding models increasingly resource schools on the basis of the students they enrol, with the explicit objective of closing the performance gap associated with disadvantage over time – making incorporation of SES in school performance measures less relevant.

One option is to combine a value-added measurement with other measures such as raw student test scores and school self-evaluation. The Victorian Department of Education and Early Childhood Development suggests that an absolute measure of attainment used alongside value-added could provide schools and teachers with the most useful information[82].

Benchmarking schools' achievement

The measures of a school's student performance outcomes outlined above (status, growth/gain and value-added) may be compared either to other schools (a relative measure) or against an absolute standard or benchmark.

As discussed above, status measures do not consider students' backgrounds, and can be perceived to be unfair. Some education systems have sought to overcome this problem by comparing groups of similar schools. For instance, Ontario

75. Council of Chief State School Officers 2005, p.5.
76. Victoria Department of Education and Early Childhood Development (DEECD) 2007, *Value-added measures for school improvement*, paper no.13, prepared by D Downs and O Vindurampulle, p.2.
77. Raudenbush 2004, p.7.
78. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2008, *Reporting and comparing school performances*, prepared by G Masters et al, pp.40-43; Victoria DEECD 2007, p.14.
79. Centre for Education Statistics and Evaluation 2014, *Value added models for NSW schools*, technical paper, prepared by L Lu and K Rickard.
80. Victoria DEECD 2007, p.5; Commonwealth DEEWR 2008, pp.40-43.
81. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2012a, p.13; citing D Ballou, W Sanders and P Wright 2004, 'Controlling for Student Background in Value-Added Assessment of Teachers', *Journal of Educational and Behavioral Statistics*, vol.29, no.1, pp.37–65.
82. Victoria DEECD 2007, pp.4-5.

compares 'statistical neighbours', or schools that are similar on multiple measures such as student population and school characteristics. In Australia, the Index of Community Socio-Educational Advantage (ICSEA) already serves a similar purpose on the *My School* website, allowing comparisons with twenty 'like' schools.

As noted above, comparing similar schools may make lower performance in disadvantaged schools appear more acceptable. Some education systems instead elect to compare schools' status measures against an absolute standard. For instance, all schools could be expected to have a proportion of their students attaining a certain score within a specified time frame. One issue with an absolute standard is that it does not provide incentives for very high achievers (who may have already surpassed this standard) or very low achievers (who may see such a standard as impossible)[83].

Growth measures may also be relative or absolute. For instance, the growth achieved by students within a school may be compared with similar schools; alternatively, schools could all be expected to attain a certain quantum of growth for their students, for instance between Years 3 and Year 5.

Finally, while value-added measures are likely to be useful for individual schools aiming to improve their performance, all value is added relative to the average school. This means that it is generally not possible for all schools to achieve positive values, or to use the measures to track changes in systemic performance over time. Alternative formulations of the value-added measure (such as measuring value-added relative to a school in a baseline year, or measuring the change in schools' value-added scores) could be used to create a benchmark of performance, but further work is required to validate these models.

Limitations of outcomes measures

Assessment of student outcomes tends to focus on academic results such as NAPLAN and PISA. This is because either full-cohort or robust sample data on these measures is regularly collected, and is capable of being compared within and across education systems. Restricting measurement of student outcomes to tests such as these may be limiting, as these tests tend to provide only a slice of a student's skills[84]. Further, international research warns of the danger of teachers 'teaching to the test' at a cost to other aspects of the curriculum[85].

Expanding the range of subjects tested or looking at results from ongoing, formative assessments such as in-school assignments[86] are two proposed methods for overcoming these issues. To look more broadly still, student outcomes could extend to school attendance; secondary school completion; employability; or social and emotional development[87]. New measures are currently being developed that seek to quantify non-academic outcomes[88].

83. Hanushek and Raymond have identified that schools closer to attaining a standard tend to change their behaviour more than those further away. E Hanushek and M Raymond 2002, 'Lessons about the design of state accountability', National Bureau of Economic Research Working Paper 10591, Harvard University, Cambridge MA, p.18.
84. M Hout and S Elliot (eds) 2011, *Incentives and test-based accountability in education*, National Academies Press, Washington D.C., p.38.
85. I de Wolf and F Janssens 2007, 'Effects and side effects of inspections and accountability in education: An overview of empirical studies', *Oxford Review of Education*, vol.33, no.3, p.382.
86. Researchers from the Bill and Melinda Gates MET (Measures of Effective Teaching) Project looked at supplemental tests as well as state tests to avoid the teachers 'teaching to the test' phenomenon. D Matthews, 2013, 'The key to evaluating teachers: Ask kids what they think', *The Washington Post*, 23 February, viewed 1 September 2014, http://www.washingtonpost.com/blogs/wonkblog/wp/2013/02/23/the-key-to-evaluating-teachers-ask-kids-what-they-think/.
87. Note these will vary between primary and secondary school. Commonwealth DEEWR 2008.
88. For instance, New South Wales recently trialled 'Tell Them From Me', a student engagement and wellbeing survey, developed by The Learning Bar. Centre for Education Statistics and Evaluation 2013, *Tell them from me student feedback survey*, viewed 1 September 2014, http://www.cese.nsw.gov.au/surveys/tellthemfromme.

Effective processes for benchmarking school performance

Education systems take different approaches to evaluating school performance. Some only utilise self-assessment; others prefer external audits or inspections; and yet others use a combination of the two.

A school that undertakes self-evaluation is understood to systematically reflect on and review the quality of educational service it provides[89]. There are advantages to self-evaluation. It is less costly than external processes, and may be both more relevant[90] and more engaging for schools. Self-assessment against set criteria or practices is useful as it promotes learning within schools by providing transparent guidance regarding areas for improvement[91], rather than simply telling schools that they must improve[92]. Despite the benefits of self-evaluation, there are also limits to such a process when it stands alone. Ultimately, self-evaluation may lack credibility[93]: it is difficult to ensure that schools are judging themselves consistently, especially if there are consequences attached to performance.

An external body performs an important role in providing this consistency in many education systems, and many systems use school inspections to fulfil this role. The OECD describes school inspection as a 'mandated, formal process of external evaluation with the aim of holding schools accountable'[94], involving external inspectors who assess schools' quality[95].

A range of educational systems have developed frameworks that accommodate both self-evaluation and external review systems, such as inspections, in recognition of their respective advantages. For example, in 2005-2006, a self-evaluation component was introduced into the UK's inspection process, after dissatisfaction with the previous model, in which the Office for Standards in Education (Ofsted) was the 'sole arbiter' of a schools' performance[96].

Some systems are electing to use peers, rather than external inspectors (employed by an Education Department or inspectorate), in the external review process. Victoria has incorporated this form of review in its school performance framework. As part of a broader drive for increased accountability and school improvement, Victorian principals will review and hold each other to account for school performance outcomes. Reviews will involve a panel of peers, including the principal of the school being reviewed, principals from other sectors, and an externally accredited reviewer. Reviews will occur at least every four years, after a school has completed internal review processes, and will consider:

- state-wide performance measures and indicators
- curriculum, assessment, reporting, teaching practices and leadership
- relationships with the wider community
- use of school resources.

Recommendations for improvement will be shared within the Department to improve accountability and sharing of best practice. Schools that fail to meet key thresholds will go through a 'priority review', which includes more in-depth analysis of the causes of the school's performance[97].

The external evaluation processes utilised by most school systems involve on-site school visits. In New South Wales, teacher accreditation provided by the Board of Studies, Teaching and Educational Standards (BOSTES) is an example of a 'desk-based' process. Teachers seeking accreditation at the Highly Accomplished Teacher level must submit a body of evidence to the BOSTES, including evidence that they have met the teaching standards, a referee report, and a report from an external observer[98]. The BOSTES convenes a Moderating and Consistency Committee that analyses this evidence and provides advice back to the Teacher Accreditation Authority, which makes the final accreditation decision[99].

89. OECD 2011a, p.435.
90. OECD 2011a, p.435.
91. Commonwealth DEEWR 2012a, pp.41-3.
92. Self-evaluations are undertaken in 21 of 32 OECD countries, two of which go on to use this self-evaluation as part of an accreditation process. OECD 2011a, p.436.
93. OECD 2011a, p.435.
94. OECD 2011a, p.434.
95. School inspections are required in 24 of the 31 OECD countries, and are a necessary part of school accreditation in seven of those. Inspections are targeted at low-performing schools in 9 countries. OECD 2011a, p.434.
96. J MacBeath 2006, 'New relationships for old inspection and self evaluation in England and Hong Kong', *International Studies in Educational Administration*, vol.34, no.2, pp.2-18.
97. Victoria Department of Education and Early Childhood Development (DEECD) 2012, *Towards Victoria as a learning community*, pp.21-22; Victoria Department of Education and Early Childhood Development (DEECD) 2013, *Professional practice and performance for improved learning: School accountability*.
98. Board of Studies, Teaching and Educational Standards 2014, *Teacher Accreditation: Apply for Highly Accomplished Teacher accreditation*, viewed 9 September 2014, http://www.nswteachers.nsw.edu.au/current-teachers/apply-for-highly-accomplished-teacher-accreditation/develop-your-application/.
99. Board of Studies, Teaching and Educational Standards 2014, *Teacher Accreditation: Submit your application online*, viewed 9 September 2014, viewed 9 September 2014, http://www.nswteachers.nsw.edu.au/current-teachers/apply-for-highly-accomplished-teacher-accreditation/submit-your-application/.

Accreditation

School accreditation sits alongside school improvement as an international trend that has come to the fore over the past two decades, within a context of increased school autonomy, decentralisation and accountability[100].

The two practices overlap conceptually, though they are distinctive, and in practice their relationship can be complex. School improvement can be described as both a goal and an ongoing process; while accreditation is often invoked as a policy lever for promoting the goal and processes of school improvement, seeking both to evaluate and recognise school quality and/or improvement.

The manner in which accreditation occurs varies considerably, and it may cover matters ranging from curriculum and assessment; access for students with disability; and facilities repair[101]; to professional development[102].

Accreditation can function as the assurance that a school has met regulatory requirements and some minimum standards. For example, in New South Wales, the BOSTES accredits non-government schools. In this instance, 'the main purpose of accreditation is to ensure that the requirements for the Record of School Achievement and/or the Higher School Certificate are being, or will be, met'[103]. In May 2014 similar processes were extended to New South Wales government schools — the BOSTES now provides the NSW Minister for Education with independent advice that government schools meet the requirements of the NSW Education Act[104].

Alternatively, school accreditation processes can operate explicitly within a school improvement context. In these cases, accreditation usually means that a school has improved the quality of teaching, learning and student outcomes, and has plans to improve further in the future. Such an approach is common in the United States, where independent accreditation bodies tend to have an improvement framework (reflecting the influence of Federal legislation, *No Child Left Behind*). The situation is quite different in Europe, where school inspections are more commonly used, and tend not to lead to accreditation. While some inspectorates do have explicit improvement agendas (the UK inspectorate's motto is 'raising standards, improving lives'), this is not universally the case. Holland's Education Inspectorate instead sees its primary role as ensuring that schools do not fall below certain standards, in relation to matters such as student outcomes, staff turnover, and school finances[105].

In Victoria, responsibility for assessing matters relevant to school quality is divided into two processes, one of which attends to minimum standards and the other of which has an improvement focus. First, the Victorian Registration and Qualifications Authority (VRQA) has responsibility for regulating education providers, ensuring that providers 'meet minimum standards and develop an integrated, quality assurance regime'[106]. These standards include:

- school governance
- enrolment
- employment of staff
- school infrastructure[107].

Secondly, the Victorian Performance and Development Culture framework aims to contribute to continuous improvement of student learning outcomes. Schools are graded at one of three levels for five elements (including induction in school, professional learning, and sources of feedback). Accreditation is obtained when schools reach the lowest of the three levels[108]. By the end of 2009, 98.4 per cent of all government schools, alongside 39 Catholic schools, had been accredited under the scheme[109].

100. A Bernasconi 2004, 'Current trends in the accreditation of K-12 schools: Cases in the United States, Australia and Canada', *The Journal of Education*, vol.185, no.3, pp.77-78; V Faubert 2009, 'School Evaluation: Current Practices in OECD Countries and a Literature Review', *OECD Education Working Papers*, No. 42, OECD Publishing, Paris, p.6; OECD 2011a, p.431.
101. J Fairman, B Peirce and W Harris 2009, 'High school accreditation in Maine: Perceptions of costs and benefits', *Penquis Superintendents' Association Research Report*, Center for Research and Evaluation, University of Maine, p.18.
102. Victoria Department of Education and Early Childhood Development, *Performance and development culture: Revised self-assessment framework*, viewed 9 September 2014, http://www.education.vic.gov.au/Documents/school/principals/management/perfdevculture.pdf
103. Board of Studies, Teaching and Educational Standards 2014, *Registration systems and member non-government schools (NSW) manual*, Sydney, p.5.
104. NSW Department of Education and Communities 2014, *Media release: school registration extended to public schools*, viewed 8 September, http://www.dec.nsw.gov.au/about-us/news-at-det/media-releases1/school-registration-extended-to-public-schools.
105. The Netherlands Inspectorate of Education 2009, *Risk-based Inspection as of 2009: Primary and secondary education*, Utrecht, http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2010/Risk-based+Inspection+as+of+2009+-+printable+version.pdf
106. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2010, *Country Background Report for Australia: OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*, p.42.
107. Victoria Registration and Qualifications Authority 2012, *Guide to the minimum standards and other requirements for school registration*, p.3.
108. Victoria DEECD, *Performance and development culture: Revised self-assessment framework*.
109. Victoria Department of Education and Early Childhood Development (DEECD), *Performance and Development Culture: Frequently asked questions*, viewed 8 September 2014, http://www.education.vic.gov.au/school/principals/management/Pages/pdculturefaq.aspx.

The Victorian model illustrates one potential tension between school improvement and accreditation processes — while the latter tends to look to standards, approving schools once they reach an identified standard or benchmark, school improvement is a continuous process, which could be (potentially) stifled by such standards. This challenge may be addressed by providing for levels of accreditation. One example of this is the Australian teacher accreditation framework, in which teachers must reach and maintain Proficient teacher accreditation; but may then apply to attain the voluntary, higher levels of Highly Accomplished or Lead teacher[110].

110. Board of Studies, Teaching and Educational Standards 2014, *How does accreditation work?* Viewed 8 September 2014, http://www.nswteachers.nsw.edu.au/future-returning-teachers/how-does-accreditation-work/.

What form should accountability take?

Accountability, within the context of education systems, refers to 'the interaction in a hierarchical relationship between those who have power and those who are delegated authority'. Accountability measures include national assessments; regulatory compliance; school inspections and school self-evaluation[111].

Accountability systems often use negative consequences (sanctions) to respond to poor performance. For example, education inspectorates in the Netherlands, England and the Czech Republic may advise their Education Ministers to impose sanctions on very low-performing schools. In Sweden, the power of the Inspectorate extends to temporarily closing schools[112]. In the UK, Ofsted conducts inspections, and identifies particular schools as 'failing' or having 'serious weaknesses'. Those schools are then placed under 'special measures', and may receive up to five monitoring inspections over 18 months[113]. If a school is still judged inadequate after this period, the Department for Education requires the local authorities to examine options in relation to that school[114].

The use of negative consequences for poor performance may lead to perverse outcomes. Jacob found that the threat of sanctions can improve test scores in low-performing schools in the short term, but this may be a result of schools 'gaming' the system. For instance, he found that, after the introduction of a comprehensive accountability policy in Chicago in 1996[115], students' performance improved in high-stakes tests without also increasing on similar, low-stakes tests, suggesting that teachers were 'teaching to the test'. He also found that there were modest increases in special education placement and grade retention[116].

De Wolf and Janssens outline a range of other potential side effects to school inspections, including 'window dressing' during inspections and teaching to the test, although they found that there was not enough empirical evidence to substantiate these claims[117]. They did, however, find evidence to support the claim that fraud increases among teachers when incentives are greater, and occurs in particular in relation to public performance indicators[118].

Accountability measures can also impact upon school morale. Elmore describes schools in the US as finding themselves 'stuck' and with no clear idea of how to improve after being put under special measures[119]. One senior manager at an Ofsted-inspected school reported that teachers were:

> *too afraid to say "come and see us and tell us what we should do" because they're not going to come down and tell you what you can do, they are going to come and close you down and put you into Special Measures*[120].

Even schools that are deemed adequate may be 'worn out' by inspections, and experience increased staff dissatisfaction, after the inspection process[121].

111. OECD 2011a, pp.430-435.
112. M Ehren et al 2013, 'Impact of school inspections on improvement of schools', *Educational Assessment, Evaluation and Accountability,* vol.25, no.1, p.19.
113. Ofsted 2014, *The framework for school inspection.*
114. Ofsted 2014, *Monitoring inspections of schools that are subject to special measures*, p.5.
115. Under this policy, students who did not meet minimum standards were held back from the next grade, and schools in which fewer than 15 per cent of students scored at or above national norms were placed on probation, with teachers at risk of losing their jobs: B Jacob 2005, 'Accountability, incentives, and behaviour: the impact of high-stakes testing in the Chicago public schools', *Journal of Public Economics* vol.89, pp.764-765.
116. Jacob acknowledges several caveats to his findings, including differences in content covered by Chicago's high- and low-stakes assessments. Jacob 2005, p.791.
117. de Wolf and Janssens 2007, p.389.
118. de Wolf and Janssens 2007, p.390.
119. Elmore 2006, p.3.
120. J Perryman 2010, 'Improvement after inspection', *Improving Schools*, vol.32, no.2, p.194.
121. Perryman 2010, p.191.

Case study: No Child Left Behind

Performance-based accountability is particularly prominent in the USA, as part of the *No Child Left Behind* (NCLB) policy. One of the main purposes of that policy is to ensure that all children have a 'fair, equal and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments'[122].

While the Federal government plays an overarching role through the targets and accountability measures in the NCLB legislation, it is the role of State Education Agencies to intervene in schools that do not make Adequate Yearly Progress (AYP) for two years in a row[123]. AYP relates to all students, with a specific focus on disadvantaged students.

Accountability measures under NCLB escalate depending on the number of years during which a school has not made AYP. They include:

- help and technical assistance, and the development of two-year plans to turn a school around
- giving students the option to transfer to a higher-performing public school in the district
- providing supplemental educational services
- replacing staff
- implementing a new curriculum
- restructuring the school[124].

It is difficult to identify the precise impact of these measures on student outcomes. A review recently conducted by the National Research Council's Committee on Incentives and Test-Based Accountability found that school-level incentives (including NCLB) produced large effects on student achievement compared to other incentives (effect sizes were 0.08) but even those were concentrated in particular subject and year groups[125]. Hanushek notes that an effect size of 0.08 may not sound significant, but that 'small gains add up', and the gains made by accountability programs surpass 'any other education program' working on a similar scale[126].

122. *No Child Left Behind Act*, 20 USC 6301 § 1001 (2001).
123. Commonwealth DEEWR 2008, pp.26-27.
124. Florida Department of Education, *Fact Sheet: NCLB and Adequate Yearly Progress*, viewed 8 September 2014, http://web.archive.org/web/20120803032049/http://www.broward.k12.fl.us/hrd/Articles/FactSheet-AYP&NCLB.pdf
125. M Hout and S Elliot (eds) 2011, pp.60-61.
126. E Hanushek 2012, 'Grinding the antitesting ax: More bias than evidence behind NRC panel's conclusions', *Education Next*, vol.12, no.2, p.4.

The evidence so far: The impact of improvement frameworks on learning

It is difficult to find evidence about the operation or the efficacy of school improvement and accreditation frameworks, for a number of reasons.

First, it is difficult to isolate the impact of a school improvement system when it has been implemented nation-wide, with no control groups, at the same time as other major policy changes[127]. In the United States, school accreditation takes place against a backdrop of Federal standard-setting (particularly, *No Child Left Behind*) and it can be difficult to disentangle the impact of state-based accreditation from that of Federal measures.

Further, the range of models, and the range of contexts in which they have been implemented, make it difficult to isolate and identify 'what works'[128]. For instance, there is some evidence available in relation to the impact of school inspections, but these inspections may occur quite separately from any explicit school improvement or accreditation agenda. Some of these studies cite 'plausible' evidence that school inspections lead to school improvement and teachers' behavioural change but findings are far from conclusive: one literature review found that school inspections had both small positive and negative effects on student outcomes[129]. The authors of that review concluded that researchers still 'do not know how school inspections drive improvement of schools and which types of approaches are most effective and cause the least unintended consequences'[130]. Findings in relation to accreditation programs are also mixed: while many teachers and leaders find accreditation to be a useful process that has enhanced the overall quality of their schools[131], others point to the stress and anxiety that can result from inspection and evaluation processes[132].

Problems arise not in the content of these frameworks, but in their implementation.

Few studies have empirically assessed the impact of school improvement frameworks on student outcomes, and those that have been undertaken present inconclusive results[133]. Studies tend to use surveys or interviews of staff rather than performing any analysis of student outcomes[134]. Repeated evaluations of the Southern Association of Colleges and Schools, which operated within a school-improvement framework, did assess student outcomes, but found no difference in school performance data (as measured by standardised achievement tests in reading and mathematics), between accredited and non-accredited schools[135].

A number of studies have found that teaching and learning are the school elements that benefit *least* from accreditation or improvement fra`meworks. Such findings have been reported in:

- Nova Scotia, where some of the lowest-scoring survey items were those relating to the impact of the School Accreditation Program on teacher practice and student achievement[136];
- New England, where interviewees 'strongly asserted' the benefits of the program for teachers and the school, but held conflicting views as to the impact of accreditation on students[137].

A similar finding was made in Queensland, where a Masters-developed tool (similar to the NSIT) was used to evaluate school performance across eight domains, over time. All Queensland schools were audited with the tool in 2010, and 25 per cent of schools were re-audited in 2011. While there were improvements across some areas after the 12-month period, the teacher practice domain showed the least improvement[138].

127. L Woessman 2006, *Efficiency and equity of European education and training policies*, CESifo Working paper No 1779 cited in Faubert 2009, p.43.
128. Faubert 2009, p.6.
129. M Ehren et al 2013, p.6.
130. M Ehren et al 2013, p.6.
131. See, eg, New England Association of Schools and Colleges 2006, *The Impact of Accreditation on the Quality of Education: Results of the Regional Accreditation and Quality of Education Survey*, NEASC 2005, p.188; Wood and Meyer 2011, p.12.
132. Fairman, Peirce and Harris 2009, , p.21; M Ehren and A Visscher 2006, 'Towards a theory on the Impact of school inspections', *British Journal of Educational Studies*, vol.54, no.1, p.53.
133. H Gaertner and H Pant 2011, 'How valid are school inspections? Problems and strategies for validating processes and results', *Studies in Educational Evaluation*, vol 37, no.2-3. An Australian paper written as part of the Smarter Schools National Partnerships also commented on the dearth of empirical evidence regarding the impact of school improvement frameworks on student outcomes: Smarter Schools National Partnerships 2010, *National collaboration project: School performance improvement frameworks*, Final Report, p.5.
134. See, eg, New England Association of Schools and Colleges p.188; Wood and Meyer 2011, p.12.
135. D Bruner and L Brantley 2004, 'Southern Association of Colleges and Schools Accreditation: Impact on Elementary Student Performance', *Education Policy Analysis Archives*, vol.12, no.34, pp.3, 12-13.
136. Wood and Meyer 2011, p.12.
137. Fairman, Peirce and Harris, p.37.
138. Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) 2012, pp.20-24.

These findings seem at odds with the evidence about the impact best practice (such as high expectations, professional development in data skills and instructional leadership) can have on teacher practices and student outcomes. This may indicate that problems arise not in the content of these frameworks, but in their implementation. Hattie's synthesis of 800 meta-analyses identifies the challenge of realising the results of any educational initiative where it matters most, finding that while professional development was likely to change teacher learning (with an effect size of 0.90), it was less likely to change teacher behaviour (0.60) and even less likely to have an impact on student learning (0.37)[139].

139. J Hattie 2009, p.120.

Enabling and sustaining school improvement through cultural change

In the view of some educational experts, accountability measures are best used as means to an end, rather than the end itself. Pressure and accountability, when divorced from support and other goals (such as development of capacity) can have a negative effect[140]. Elmore suggests that this may be because teachers are already operating 'more or less at the limit of their knowledge and pedagogical skill', and adding pressure, without also providing support or guidance as to how to reach goals, may have little impact[141]. Fullan similarly argues that using test results alone to punish or reward schools 'assumes that educators will respond to these prods by putting in the effort to make the necessary changes … it assumes that educators have the capacity or will be motivated to develop the skills and competencies to get better results'[142]. Drivers of school improvement are far more likely to be successful if they foster intrinsic motivation; engage educators and students; inspire team work; and affect all teachers and students[143].

There is some support for this in the McKinsey study, which found that teachers in successful systems received 56 per cent of all support initiatives, but only 3 per cent of accountability measures, such as teacher appraisals. Teachers in these systems were held accountable through their students' learning and collaborative practice with their peers:

> *By developing a shared concept of what good practice looks like, and basing it on a fact-based inquiry into what works best to help students learn, teachers hold each other accountable to adhering to those accepted practices*[144].

A number of education systems provide support for their schools as part of their improvement or accreditation frameworks. For instance, in the Northern Territory, a coaching model is used to develop principals' skills[145]; in South Australia, the Department of Education and Children's Services is responsible for developing workforce capabilities and system capacity as part of its Improvement and Accountability Framework[146]; and in Victoria, the Performance and Development Culture framework encourages effective induction and mentoring support for teachers[147].

Cultural change of the type described by Fullan may be difficult to achieve, but it *is* possible. Moreover, it is *essential* to sustaining educational improvements. Tucker observes that a 'sustained emphasis on education quality … carries enormous implications' in terms of garnering support at all levels, from government to educators and the broader community[148].

The McKinsey study identified 13 'sustained improvers' – systems with at least five years of consistent rises in student performance across multiple data points and subjects. As the study reports:

> *For a system's improvement journey to be sustained over the long term, the improvements have to be integrated into the very fabric of the system pedagogy*[149].

140. R Elmore 2007, *Educational Improvement in Victoria*, p.2; M Fullan 2011, 'Choosing the wrong drivers for whole system reform', *Centre for Strategic Education Seminar Series*, Paper no. 204.
141. R Elmore 2006, 'OECD activity on improving school leadership' Paper presented at the International Perspectives on School Leadership for Systemic Improvement International Conference, Harvard University, 6 July.
142. Fullan 2011, p.8.
143. Fullan 2011, p.3.
144. Mourshed, Chijioke and Barber 2010, p.75.
145. Northern Territory Government, *School accountability and performance improvement framework*, viewed September 8 2014, http://www.education.nt.gov.au/__data/assets/pdf_file/0005/15773/SchoolAPIF.pdf
146. Government of South Australia, *DECS improvement and accountability framework,* viewed 16 July 2014, http://www.decd.sa.gov.au/docs/documents/1/DecsImprovementandAccou-1.pdf
147. Department of Education and Early Childhood Development, *Performance and development culture: Revised self-assessment framework*, viewed 16 July 2014, http://www.education.vic.gov.au/Documents/school/principals/management/perfdevculture.pdf
148. Tucker 2012, p.40.
149. Mourshed, Chijioke and Barber 2010, p.72.

GPO Box 33
Sydney NSW 2001
Australia
T 9561 1211
www.cese.nsw.gov.au

© October 2014
NSW Department of Education and Communities

NSW GOVERNMENT Education & Communities