Building Capacity in a Self-Managing Schooling System:

The New Zealand Experience

Viviane Robinson

Stuart McNaughton

and

Helen Timperley

The University of Auckland

## Building Capacity in a Self-Managing Schooling System:

## The New Zealand Experience

The recently released 2009 PISA results present a good news/bad news picture of achievement for New Zealand's 15 year olds (OECD, 2010b). The good news is that, on average, New Zealand 15 year olds achieve at significantly higher levels than the OECD average. For reading literacy, only two OECD countries and two newly participating countries, Shanghai-China and Hong Kong-China, had significantly higher scores. These results are consistent with those obtained over at least a decade of international comparisons (Elley, 2005).

There is bad news in the results too. The variance in reading achievement in the PISA data is 22% higher than the OECD average. Due to the comprehensive nature of most New Zealand schools, about 95% of the variance in achievement lies within schools, meaning that there are pockets of very low achievement in most New Zealand schools. Perhaps of greatest concern is the strong relationship between socio-economic background, ethnicity and achievement. Across the 34 participating OECD countries, 14% of the variance in achievement was explained by socio-economic background, while for New Zealand the figure stands at 17%. This is in stark contrast to Canada and Japan where the comparable figure is just 9%. New Zealand stands with Belgium in having high average performance and large socioeconomic inequalities (OECD, 2010a). Together, these data support the OECD conclusion that the New Zealand system can be characterized, at least at secondary level, as one of high performance and low equity.

Māori students (from the indigenous community) and Pasifika students (from both recent immigrant and longer established Pacific Island communities) are over represented in the tail end of the achievement distributions. Currently, Māori make up 22% of the school age population and Pasifika groups a further 9.6%. Projections are that these two groups will make up 46% of the school population by 2026 (New Zealand Ministry of Education, 2010).

We have known about the gaps between Māori and more recently Pasifka children, and other children, for decades with little evidence of wide spread change (Lai *et al*., 2009). This seemingly intractable pattern of high achievement but low equity is likely due to both school-related and out of school factors or "OSF's" (Berliner, 2009). With the more explicit recognition of the significance of OSFs there has been a growing sense of pessimism about

how much teachers and researchers can do just focusing on the school system's effectiveness. Large scale system effectiveness is substantially constrained or reversed by OSFs such as disparities in incomes, health, housing, and neighborhood infrastructure (Berliner, 2009; Wilkinson and Pickett, 2009).

We do not assume that the only responsibility for improvement lies with educators. Just as there are multiple causes, there are multiple possible points of intervention to increase the effectiveness of schools. Optimally, both changes to school systems and to the lived conditions of communities may be needed to produce lasting equitable changes. But the variability between countries, at any one time and over time, suggests that systems can be improved (Levin, 2008; OECD, 2010c). Similarly, the evidence from targeted interventions (McNaughton and Lai, 2010) suggests that it is possible to make improvements in school systems.

In this paper we focus on the conditions required for effective and equitable system-wide educational interventions. By effective, we mean that outcomes for Māori students, Pasifika students and students in poor schools match nationally expected distributions of achievement. In this paper, our primary focus is on literacy and mathematics outcomes because they are core contributors to academic achievement and there are long term costs of low achievement in these subjects to both an individual (e.g., in well being, health, further study, employment, and income) and to society as a whole (e.g., in family well being, social cohesiveness, and economic output).

We begin by using evidence from analyses of systems that are either high performers or successful reformers to establish a broad framework against which we evaluate two recent attempts to build the capacity of New Zealand schools to lift the excellence and equity of student achievement. In making this evaluation, we are attentive to context, particularly the nature of school self management in New Zealand and to the interactions and tensions that exist between various system characteristics. Finally, we discuss what the two cases reveal about the challenges and limitations of building system-wide capacity in New Zealand's self-managing context.

## The Origins of Our Evaluative Framework

We used two different approaches to constructing an evidence-based framework for evaluating New Zealand's recent school improvement strategies. The first approach, based on

comparative system performance, drew on an analysis of the characteristics of high performing systems in the recent 2009 PISA survey of the achievement of 15 year olds in reading, math and science (OECD, 2010b). High performing systems were those which had "high participation, high quality, high equity and high efficiency" (OECD, 2010c, p. 231). Of the 34 OECD countries who participated in the 2009 PISA study, those that best met these five criteria were Canada, Japan, China–Shanghai, China–Hong Kong, and Finland.

While this first approach yields some *possible* lessons for New Zealand, there are two important caveats. One is that the factors associated with high system performance are not necessarily the same as those associated with system improvement (Sleegers *et al*., 2002). This is important, as our paper provides a critical analysis of two New Zealand reform initiatives rather than a critique of overall system performance. In addition, the success of high performing systems may depend on historical and cultural factors, which, though not directly measured by the OECD surveys, provide the critical explanations of the correlations between the measured system characteristics and education performance.

A critical feature of the New Zealand context is that 22% of the current school age population identify as Māori, and for them and their communities, a single focus on achievement, will not gain their trust of the schooling system. That is why recent New Zealand education policy has emphasized a second equally important culturally specific outcome – success as Māori (New Zealand Ministry of Education, 2008). Whatever lessons about raising achievement can be derived from the OECD work, they have to be adapted to a context in which both these outcomes must be achieved. A second important feature of the NZ system is that each school is self-managing (New Zealand Government, 1989; 2001). It is governed by its own board of parent-elected representatives who, among other responsibilities, appoint and appraise the school's principal. The Board, which includes the school principal, has discretion to govern the school as it sees fit, subject only to the law. School boards must report to the ministry and to their communities about student achievement and are reviewed approximately every three years by the national Education Review Office.

The second approach we used in developing our evaluative framework was based on relative system improvement rather than relative system performance. The OECD reports referred to earlier provided some relevant evidence, because they compared the 2006 and 2009 PISA results to identify successful reformers, and described aspects of the reform process in those countries. Our main source of evidence about successful reform, however, was the 2010

McKinsey report entitled "How the world's most improved school systems keep getting better" (Mourshed *et al*., 2010). The report used the results of multiple national and international assessments to construct a standardized achievement scale which was used to classify system performance as poor, fair, good, great and excellent. The authors then developed criteria to identify the systems that had made significant, long-lasting and widespread gains in performance either within or across these five categories (Mourshed, *et al*., 2010). Once countries that had made sustained improvement had been identified, the authors then identified both those reform strategies that were specific to particular stages of development and those that were more universal, but employed in context-specific ways.

According to the McKinsey analysis, the New Zealand system is on the boundary between the fair and good categories. This is somewhat lower than suggested by the PISA data discussed in the earlier section − a difference that is probably due to the fact that system performance in the McKinsey report is based on New Zealand students' performance over many years, on multiple international surveys in reading, maths, and science at several different age levels. Given that New Zealand's performance sits on the boundary between fair and good, we have paid particular attention to the reform strategies of countries that have moved from the good to great category of student achievement. These countries are England, Canada (Ontario), Hong Kong, Singapore, and South Korea. There is a considerable overlap between these countries and those we earlier identified as high performers from the PISA study.

In finalizing our evaluative criteria, we have integrated insights from both our high performers and successful reformers analyses. We also draw on additional research examples and evidence to expand or illustrate the criteria in action (Bryk and Schneider, 2002; Levin and Fullan, 2008; Odden, 2009).

## Systemic Conditions Associated with Improvement

In the remainder of this section we outline six conditions that are associated with successful system level reform in countries that have improved to the level to which New Zealand now aspires. There is wide variation in how these six conditions are manifest in the improving systems, because reform strategies interact with each other and with the context in which they are implemented.

**System-wide Commitment to Educational Improvement**

In successful systems, the importance of education for individual and collective well-being is almost universally accepted. Political and professional leaders have convinced the public that relative to other demands on the tax dollar, a high priority should be given to education spending. The commitment is to high levels of performance for all students – not just a selected few – supported by the belief that with the right support from parents and teachers, and appropriate effort from the students themselves, all students can succeed on intellectually challenging curricula. Belief in the importance of effort and high quality support has displaced the view that educational outcomes are largely determined by inherited talent.

Political and professional leaders have made these commitments concrete and public in the form of ambitious and inclusive goals. In Ontario, for example, the new government developed targets for increased high school completion and improved achievement in Grade 6 reading, writing and math. These goals were endorsed by all stakeholders, pursued by carefully planned strategies and generously resourced (Levin, 2008). In 2010, the Japanese government announced a ten year growth strategy that included targets for reducing the proportion of low achieving students and increasing the proportion of students reporting an interest in reading, math and science to the level of the OECD average (OECD, 2010c).

Public commitment to the achievement of specific goals or targets poses political risks, as falling short will bring criticism from the media and opposing parties. On the other hand, there is also risk for political leaders in not responding strongly and publicly to evidence or events that suggests the education system is in trouble. In Ontario, the catalyst for reform was four years of industrial disruption that had many families exiting the public school system (Mourshed *et al*., 2010). The risks are mitigated by courageous leadership which takes the public and stakeholder groups into its confidence about the nature of the problem and invites them to participate in the process of resolving it (Odden, 2009).

**Ambitious Universal Standards**

In improving systems, the belief that all students can and should learn is made concrete through the formulation of standards that describe what all children at different ages should know and be able to do. Universal application of core standards means that all students, rather than just an elite group, have access to a cognitively demanding curriculum.

The standards articulate a learning progression that reduces duplication of teaching across year levels, provides clear benchmarks of a child's progress and guides the development of tightly aligned resources and texts. Most successful systems assess performance against the standards using sophisticated written external exams that provide valid measures of higher order knowledge and thinking skills. There is little reliance in high performing systems on computer-scored multiple choice assessments. While the detail of how standards are formulated, and how far they have shaped classroom practice varies widely, they "shape high performing education systems by establishing rigorous, focused and coherent content at all grade levels; reducing overlap in curricula across grades; reducing variation in implemented curricula across classrooms; facilitating coordination of various policy drivers, ranging from curricula to teacher training; and reducing inequity in curricula across socio-economic groups" (OECD, 2010c, p. 235).

It is true that in some very high performing systems, such as Finland, there is little reliance on centrally developed standards and assessments. The professional capacity of the teaching force and the cultural capital of the community are so high that the requisite knowledge is largely located in, and developed through, the collective cognition of teachers. In New Zealand, a more central approach is needed until a teaching profession can be established that is more selective and more highly qualified than at present.

**Developing Capacity at Point of Delivery**

Ambitious universal standards will not work unless leaders and teachers have the capacity to deliver them. Across systems that are making the shift from good to great performers, the focus of reform efforts is the professionalization of teaching and leadership through embedding routines for teaching and leadership excellence. There are also substantial efforts to raise the status of the profession by making it more selective and providing more professional working conditions and better remuneration.

Reform of initial teacher education has involved shifting it from lower status teacher training institutions into higher status universities, incorporating rigorous training in early and accurate diagnosis of student learning needs, application of research-based teaching practice, deep knowledge of academic subjects and their associated pedagogies, and rigorous inquiry into practice.

In the high performing countries of East Asia, a tradition of collective lesson planning and lesson study, led by master teachers, provides teachers with high quality, job-embedded opportunities to learn from expert peers. In systems as different as Japan, Finland and Shanghai-China continuous improvement at the school level is pursued through teacher cultures in which the "practice of individual teachers is open to inspection by the other teachers in the school, and the quality of teachers' practice is seen as a matter for all teachers to be concerned about" (OECD, 2010c, p. 244). Routines that drive improvement are not limited to an annual appraisal approach, or siloed in a project or initiative. They are a regular feature of every teacher's work, made possible in some cases, by a trade-off between class size and blocks of time for collaborative teacher learning.

Efforts to build leadership capability echo the research evidence about how leaders create the conditions for improved learning (Robinson *et al.*, 2008). In high performing systems, there is a strong emphasis on helping leaders make the shift from administrative to instructional leadership.

**Professional Forms of Accountability**

Although there is wide variation across high performing and improving systems in accountability policies, the common theme is a shift from more bureaucratic top-down forms to more emphasis on accountability to internalized professional norms, to peers and to parents and students. This shift happens hand in hand with the development of a more professional teaching workforce, that is, one with higher knowledge and skills, higher status and more consistent use of effective practice.

In high performing East Asian systems, the tradition of shared lesson planning and lesson study, led by an expert master teacher, instils a strong sense of collegial accountability. A strong collective focus on student learning is not enforced by rules, but by mutually felt obligations to standards of instruction and learning (Kruse *et al.*, 1994).

The OECD reports are careful to point out the relationship between professional accountability and teacher capacity. As that capacity increases, leaders and teachers are given more discretion about how to teach and manage resources in ways that deliver the expected results. While there is rigorous monitoring of results against system-wide standards, high performing systems use the results to diagnose and address teacher and student learning needs, rather than to reward or sanction schools and teachers. In other words, robust

performance assessment is common to both bureaucratic and professional forms of organization – it is the use of the performance data that differs under each approach.

There is a similar type of interaction between accountability and school autonomy. In systems with public accountability for performance on standards, higher levels of school discretion are associated with higher performance. In systems without such accountabilities, higher levels of autonomy are associated with lower student achievement.

**Strategic Resourcing**

The total amount spent on education, whether calculated by percent GDP or on a per capita basis, is not predictive of system performance. The McKinsey report shows that systems with similar levels of per pupil expenditure achieve widely varying student results (Mourshed *et al*., 2010, p. 21). New Zealand per pupil public expenditure is in the same band as that of Estonia, South Korea, Hungary and the Czech Republic all of whom outperform it on the standardized McKinsey achievement scale (Mourshed *et al*., 2010, p. 21). More important than total spend are the strategic choices that are made about how to spend the money and the implication of those choices for the excellence and equity of system performance. In high performing systems, a greater proportion of funding is allocated to the core business of teaching and teacher learning than in many lower performing systems. Equity is pursued through a range of different strategies to ensure that funding and teacher expertise are directed to students from poor, remote or immigrant communities. Where there is central control over the teacher workforce, expert teachers and leaders may be moved, on either a temporary or permanent basis, to underperforming schools as part of a comprehensive effort to ensure equitable learning opportunities.

**Institutionalizing the Improvement of Practice**

The literature on educational change contrasts two different approaches. Under the "planned change" approach, educational change results from detailed central planning of innovation, which is then adopted by implementing agents, who are carefully managed to ensure consistency of implementation. Teachers are the recipients rather than origins of improved practice. In the second "cultural" perspective on change, a more incremental, local, and teacher-driven approach is taken where the emphasis is on creating the conditions, such as professional learning communities, that enable teachers to improve practice in their own school context (Sleegers *et al*., 2002).

Different combinations of these approaches are evident in high performing and improving systems, reflecting differing traditions of public sector reform and the differing capacity and professionalism of the teacher workforce. Regardless of the predominant approach, high performing systems have systematic institutional routines for the improvement of practice at both system and school level. At system level, there is a clearly identified agency or agencies that are responsible for the progress and performance of the system as a whole.

Considerable attention is given to the coherence of the reform effort and to the consistency of its implementation. System coherence, consistency and monitoring of progress all require sophisticated data gathering, analysis and reporting capacity. One of the reform strategies that is common to systems that have moved from fair to good performance, is substantial investment in the infrastructure required to support these data functions.

In Singapore, which is arguably the paradigm example of the planned change approach, reform begins with a national and international search for best practice, followed by extensive trials of new practice in carefully selected schools. The trials provide rich feedback to the developers – usually teams of policymakers and practitioners – about how to revise the policy, and its associated resources and training opportunities. There is enormous attention to the details of implementation, with no policy announced without an accompanying plan for building the capacity to implement it (OECD, 2010c).

Policy, organizational and teacher learning is fostered in high performing and improving systems by structures and cultures that include ongoing and job-embedded routines for continuous learning and improvement. There is, in the words of Richard Elmore, a standard practice of improvement (Elmore, 2004). Political and professional leaders take responsibility for creating the focus and rationale for educational improvement and for the infrastructure which provides the analysis and feedback which enables detailed monitoring of its progress throughout the system. At the school level, there is a varying balance across systems between supporting teachers in the implementation of centrally designed change and in supporting their collective development of improvement. In either case, the centre takes responsibility for developing or locating the teacher and leadership capacity which improvement requires.

### Two New Zealand Approaches to Building School Capacity

New Zealand has taken a two-pronged approach to school improvement (Annan, 2006). The first, which lasted from the early 90s until 2009, targeted schools serving low income

families, which were experiencing problems with aspects of school governance, management or student achievement. The second approach involved the introduction of policies designed to lift the performance of the whole system. Our second example, the introduction of national standards, is the most recent example of such policy.

**School Improvement through Partnerships with School Clusters**

The relatively extreme form of self-management introduced in 1989 left the Ministry with limited powers to intervene in schools about which they had concerns (New Zealand Government, 1989). By the mid 1990s it was apparent that some schools, particularly those serving low socio-economic communities, were struggling (New Zealand Education Review Office, 1996). Over the following few years, a New Zealand schooling improvement (SI) model evolved under which selected schools, usually organised into geographical clusters, received extra resources in return for engaging in needs-based school improvement activities. By 2006, 15-20% of the nations' students attended schools in one of 19 schooling improvement clusters. Early clusters focused primarily on the improvement of board governance, school leadership and school-community relationships. Later clusters were required to focus much more directly on the analysis and improvement of student achievement. It is the work of these later clusters that is the subject of this section.

The criteria for school participation in SI were that the school served students from low SES communities who were achieving below an acceptable level. There were a number of avenues by which schools entered SI: some schools put themselves forward, others were nominated by the ERO, and the Ministry of Education also identified possible schools from a set of indicators based on school enrolment, staffing trends and financial performance. Since there is no national testing in New Zealand until Year 11, these decisions were frequently made in the absence of accurate and comparable information about school performance. As a consequence, the Ministry required an achievement analysis to be completed, usually by a third party, at the beginning of the SI process. Those schools with achievement profiles that were substantially different from the national profile were grouped together in geographically based SI clusters. The next step involved working with an external coordinator to develop an agreed theory for improvement and an action plan. Once these activities were completed, the cluster was provided with additional funding for a cluster coordinator and professional development for leaders and teachers. An officer from the Ministry of Education provided support and oversight of the cluster. Typically these roles were held by different people.

Providing a definitive answer on the extent to which student achievement improved under cluster-based SI is difficult. Over the period of the evaluation, the team helped the clusters to collect high quality data. By 2009, 13 of 19 clusters provided some data that allowed for analysis at two points in time ($n$=15,478 students in 99 schools). Some clusters had strong moderation procedures in place, in others the cluster informants were not able to describe the procedures used. For the areas of reading comprehension (Years 3-9), only six clusters could provide data for all year levels but the overall effect size for all gains compared with nationally expected gains was $d$=0.29 . In writing (Years 4 − 9) four clusters provided some data at some year levels and overall the effect size gain compared with nationally expected gains was greater at $d$=0.52 (Lai, McNaughton & Hsiao, 2010).

In the remainder of this section we use the six criteria for developing system-wide capacity to evaluate this more recent phase of New Zealand's school improvement work. The evidence on which the section is based was collected during a three year formative evaluation of all New Zealand's school improvement clusters. The second and third authors took a leadership role in this evaluation the results of which are available in a series of reports (Timperley *et al*., 2008; Timperley *et al*., 2010; Lai *et al*., 2010; Parr *et al*., 2010).

*System-wide commitment to educational improvement*

In its SI work, the Ministry of Education walked a fine line between central regulation and school autonomy. For example, it required all schools to focus on achievement and to use national and not just local benchmarks to assess student achievement problems and the clusters' progress in addressing them. These requirements were one expression of its commitment to a more equitable and effective schooling system. On the other hand, the Ministry of Education gave schools substantial control over their choice of improvement strategies. Although each cluster was required to develop its own "theory of improvement' which would show how the student learning needs that had been revealed by the achievement analysis and other evidence would be addressed, there was in practice, no requirement to develop an effective theory.

Interviews of leaders and teachers showed that the process of engaging schools in the achievement analysis and in writing the theories of improvement led to high levels of understanding about the central purpose and aspirations of the intervention. Once we move beyond the shared acknowledgment of the problem and the aspiration for improvement, a different story emerges. The structures and culture of school self-management limited the

Ministry of Education's ability to ensure that this general commitment was translated into effective school level action. Although nearly all schools with identified achievement problems did agree to participate in a cluster, there was, in many cases, a substantial gap between participation and serious engagement in the clusters' activities. Interviews of cluster coordinators and professional development providers included frequent references to the constraints that volunteer participation placed on their work. Their fear that schools would withdraw from the work, substantially limited the degree of challenge and urgency that coordinators, facilitators and ministry officials communicated to school leaders.

Skilful oversight of each cluster by officers from the Ministry could have ensured more consistent school commitment to the improvement work. However, these external facilitators and providers received limited guidance from the centre and varied widely in their understanding of the extent to which their role involved leadership, coordination, facilitation or support. This problem was exacerbated by the location of these people throughout national and district Ministry offices and by their multiple other responsibilities. All of these factors produced a situation akin to every school cluster "developing its own drugs" (OECD, 2010c, p. 244). Some of those drugs were more effective than others.

*Developing capacity at the point of delivery*

Analysis of the funding provision agreements showed that the intention to develop capacity among teachers was common to all cluster interventions. In reality, however, there was lack of clarity about how the focus for cluster professional development was decided and wide variation in both leaders' and teachers' participation in the professional development activities of each cluster (Timperley *et al*., 2008). Although most teachers understood that the professional development was designed to address the achievement problems in the schools, they had not made the connection between that purpose and the need to change their own practice (Le Fevre, 2010).

Developing capacity at the point of delivery must also involve leaders if that capacity is to become more than patches of brilliance in particular classrooms. Interviews of school leaders and an analysis of cluster development plans showed that professional development for school leaders was relatively rare. Most improvement plans had been written by school leaders and yet they themselves had not seen the need to learn more about how to lead the achievement of the students for whom they were responsible.

School leadership capability issues were exacerbated by the New Zealand system for appointing and appraising principals which gives responsibilities to each school board. Previous research has shown that boards in low socioeconomic communities have limited capacity to identify the kinds of curriculum and instructional leadership needed to reform schools (Robinson *et al.*, 2003). In addition, access to leadership capability is constrained by a remuneration system in which principal salary is largely determined by school size. Since the roll of SI schools is often small and declining, the result is often that the least experienced principals can be appointed to the most challenging schools.

*Professional forms of accountability*

The OECD and McKinsey analyses both emphasized the interaction between the level of professional capability and professional autonomy. As teachers develop capability they develop a better understanding of the norms and practices that are required if school improvement goals are to be met. Such understandings enable them to be collectively accountable and responsible for the quality of leadership, teaching and learning. Given the mismatch between the initial capability of staff in cluster schools and that required to solve the achievement problems, it seems likely that the New Zealand school improvement approach erred too much on the side of professional rather than central accountability.

Although clusters were required to submit plans for additional funding to the Ministry of Education and provide milestone reports identifying how the funding was spent, in reality no plans were rejected and funding was rarely withheld. In addition, the centre did not sufficiently tailor its accountability approach to the demonstrated capabilities of each cluster. For example, while all clusters had access to aggregated cluster data, some did not agree that identifiable school-level data would be available to all school leaders within the cluster. This made it very difficult for schools to either learn from one another or to hold each other to account for non-implementation of plans or lack of success. In other clusters, where data on both teaching practice and student achievement were identified by school, there was not only a strong sense of accountability but also a strong focus on mutual learning from the information provided.

*Strategic resourcing*

The intention of government was certainly to increase the equity of outcomes by providing targeted support to schools serving low income communities. SI contracts directed additional

funding to improving teaching through better school-based infrastructure, such as data management systems, and professional development. The degree of autonomy afforded to each cluster meant, however, that the investment was often not used strategically. Sometimes schools developed their theories of improvement without reference to research on the impact of various types of professional development interventions. Facilitators, professional developers and ministry officials who supported cluster activities had highly variable knowledge of school reform and of effective instructional and leadership practice. In addition, some of those within the Ministry responsible for school improvement clusters believed they did not have the authority needed to ensure that cluster decisions were based on evidence about what worked.

*Institutionalizing the improvement of practice*

Predictably, this approach to schooling improvement allowed for a few highly effective interventions to develop, particularly where external experts who engaged with the clusters were knowledgeable about improvement strategies. The majority of clusters, however, struggled to make the same rate of progress. Unfortunately, the emphasis on local solutions meant there was no mechanism to institutionalize good facilitation, leadership or teaching practice across the clusters. The result was highly variable practice with little systemic understanding of or responsibility for the variance. The point can be illustrated with an example about the analysis of achievement data.

In most clusters new data were collected each year to judge progress and decide on new priorities. In one well-established cluster, a meeting to discuss the writing data was observed by researchers. Those present understood that progress had been minimal over the previous three years despite writing being a professional development focus. Despite that, the cluster group decided to continue with the same professional development approach because they believed that teachers needed more time to implement what they had been taught. No consideration was given to the adequacy of the professional development or to collecting classroom observations to test exactly how the teachers were attempting to teach writing. The members of this cluster needed better help with both their understanding of writing and their skill in identifying and testing their assumptions about how to improve.

Although both these capabilities were present in a second cluster that was working with a university research partner, there was no structural mechanism in the NZ schooling improvement approach for identifying and sharing this expertise. In this second cluster, the

university researcher was able to guide the schools through successive diagnoses of student and teaching needs in reading comprehension. Together, the researcher, cluster leaders and teachers developed increasingly sophisticated intervention strategies to the point where the student reading profiles matched national profiles for students of the same age.

Achieving a systemic response to SI within a self-managing education system is challenging. As long as the aim is to retain schools as self-managing organizations rather than to differentiate eligibility for self-management on the basis of capacity, it is almost inevitable that schooling improvement will have a localized flavor with highly variable results. This is not to say there were no standardized practices. Every cluster completed an achievement analysis and wrote a theory for improvement. Every cluster tried to develop capacity at the point of delivery.

Beyond these commonalities, however, there was wide variation in both the nature and effectiveness of SI strategies and little evidence-informed feedback from which cluster leaders and facilitators could learn. Where strong external expertise was available to clusters, local initiatives were soundly based with strong collegial accountabilities evident. Some schools developed profiles of achievement that matched the national profile. Other local efforts were more idiosyncratic and less soundly based. The limited initial capacity of some schools could have been have been offset by the presence of a strong systemic capability to work with schools to develop and implement interventions that were both evidence-based and responsive to the local context. Such central capability could have also provided systematic opportunities to develop strong collegial accountability, both within and across clusters.

**System Reform through the Introduction of National Standards**

In 2009 the centre right party contesting the New Zealand national elections proposed a plan to ensure greater educational success at school by implementing national standards in reading, writing and mathematics. The proposal stressed reducing the 'tail of achievement' by setting standards and ensuring accountability to parents by giving them "straight answers to straight questions" about their children's achievement (National Organisation, 2008).

The proposal became an official policy of the Ministry of Education when the National Party won the election in 200?…The policy stated that the National Standards (NS) would enable better student learning and progress by providing teachers, parents, and students a clear understanding of expected levels of achievement. The framework that encompasses the

standards themselves explicitly refers to using the standards to reduce within-school disparities in achievement, create positive outcomes for Māori (in cultural identity terms as well as achievement) and prevent low achievement caused by low expectations based on ethnicity or socio economic status (New Zealand Ministry of Education, 2009a, pp.6-7). The Ministry of Education (2007) already had developed systematic descriptions of learning progressions (essentially rich benchmarks) in literacy and numeracy and the NS operationalised these benchmarks.

The National Administration Guidelines (NAGs), which are the mandated requirements for schools, were revised in October 2009 (New Zealand Ministry of Education, 2009c) to require all schools in Years 1-8 (5 year olds to 13 year olds) to report to students and their parents on progress and achievement against the standards. Schools are required to report to parents in writing twice a year in accessible "plain language" (New Zealand Ministry of Education, 2009b). School-level data are also required on numbers of students at, above, below, or well below the standards, and the school data are to be disaggregated by gender and Māori and Pasifika ethnicity. Schools report the data to the Ministry are through annual Board of Trustees reports.

Since there are no mandated tests at primary level in New Zealand, schools are expected to draw on a variety of available assessment tools to obtain evidence of their students' progress and achievement. Teachers then use what has been called "overall teacher judgment" to evaluate each student's performance in relation to the relevant national standard (New Zealand Ministry of Education, 2009a, b). The first annual report under the new regulations is set to be completed by May 2011, which provides an 18 month time frame from no standards to full implementation. The 2010 year was set aside for further refinement of the standards and some professional development. Despite substantial objections from the sector, and some collective resistance, schools were told that NS was to be implemented fully at Years 1-8 across the system within the time frame and without any "trial" or graduated roll out.

At least three theories of action are entailed in this development. One is that greater shared explicit understanding of expected levels will enable teachers to teach more effectively. The second is that being accountable to parents is a major lever for system change. The third is that the system has or can quickly develop, the capability and capacity to implement the standards effectively.

*System-wide commitment to educational improvement*

The criterion of system wide commitment has not yet been met. This reflects tensions between self governing schools and national policy requirements. There is a lack of clarity around what the policy framework is designed to improve and, more specifically, how the NS might relate to solving the achievement disparities. The responses of educators to the consultation process indicated unclarity about how the standards (only set for Years 1-8) might relate to progress throughout the curriculum and in senior secondary qualifications (Wylie *et al*., 2009). Similarly, while schools are expected to set appropriate targets for their particular students (New Zealand Ministry of Education, 2009a), the Ministry is explicit in not setting a blanket target for schools (New Zealand Ministry of Education, 2009b) requiring only that the school level data indicate school strengths and identified areas for improvement, the basis for identifying areas for improvement, and planned actions for lifting achievement (New Zealand Ministry of Education, 2009c). The lack of specification about school level reporting has confused school leaders and compromised the possibility of consistent reporting to government about school performance. If equity goals are to be pursued by individual schools, school clusters and by policy makers themselves, common indicators of progress are needed across the system. Experience in successful school improvement clusters has shown that a very detailed evidence base about learning and teaching needs is critical to teachers and leaders learning how to raise the achievement of Māori and Pasifika students (Lai and McNaughton, 2009; Timperley and Parr, 2009). So far the reporting requirements are insufficiently specified to support such learning.

*Ambitious universal standards*

The criterion of ambitious universal standards is partially met in the policy. The standards are ambitious, in the sense that they are pegged to valid measures of higher order skills. The progressions have been built from an amalgam of professional judgments, expert identification and the use of national assessment data. Indeed, the literacy progressions and their associated standards have been criticized as "aspirational" (Wylie *et al*., 2009), and there is some evidence from developmental descriptions of writing that that is the case. It is important to recognize that there is no sense in the standards that target groups, such as Māori and Pasifika students would not be expected also to reach these.

Currently there are no standards for the secondary years and the international evidence suggests this dislocation will reduce effectiveness across the system (Levin, 2008). Although

there is ongoing work to produce standards for the secondary years, it is not clear when this will be completed.

*Developing capacity at the point of delivery*

Considerable responsibility for successful implementation is devolved to the leadership and teachers in a school. External support was provided in 2009 and 2010 in the form of one-off workshops and by integrating national standards seminars into existing national professional development contracts. Resources also were provided including a self review tool which teachers, principals and Boards could use to identify their areas of strength and needs in relation to implementation of the standards.

By explicitly requiring teachers to use a variety of assessment to make overall judgments about how students met the standards, the policy requires teachers to act as 'adaptive experts' making complex judgments from multiple sources of information rather than to act as 'routine experts' who follow existing templates (Darling-Hammond and Bransford, 2005). This approach reflects a long history of teachers acting relatively autonomously and of the profession itself generating innovative practice (McNaughton, 2002). But the need to produce dependable national data about the achievement of complex higher-order standards, requires consistency in both within-school and between-school judgments. The evidence shows that being able to produce such judgments, especially in newer assessments in writing, requires months of moderation training within and across a cluster of schools (Parr and Timperley, 2010). It also shows that reporting to parents will require extensive professional development because capability is at low levels and there is considerable variability across schools (Koefoed, 2009). Currently, there is no national system that sets out the requirements for the moderation and reporting of teacher judgments. Without substantially more investment in the development of a coherent and robust moderation process, dependable and consistent teacher judgments are unlikely (Raphael *et al*., 2009).

*Professional forms of accountability and strategic resourcing*

In theory, shared clear standards are an important tool in developing professional accountability. As noted above there is national evidence from a programme of professional development associated with gains for Māori and Pasifika students (Timperley and Parr, 2009) as well experimental interventions with schools serving those students (McNaughton and Lai, 2009) for the significance of professional forms of accountability. But these pockets

within the system also demonstrate that these dimensions of professional learning communities need to be systematically developed.

The tensions in the national commitment to developing this capability are also illustrated in resourcing for implementing the NS. Some needs were resourced as we have noted (such as self review tools). But in other areas the resourcing has not been as strategic as it needs to be. For example, the Ministry of Education has not directly required nor resourced a national system-wide student data management system (SMS) which could deliver the aggregated school level data. Schools have bought and designed their own data management systems. The previous school improvement example referred to evidence, gathered in 2009, of highly variable capability in data management such that few school clusters could report accurate and reliable longitudinal data. Unless capability is either developed or in every school or centrally provided, these difficulties will jeopardize the reporting and learning goals of NS.

*Institutionalizing the improvement of practice*

This criterion is not met in the NS policy and its roll out. We have noted two features already: the restricted timeline and the deliberate rejection of a roll-out design. Neither the "planned change" approach nor the "cultural" change approach was used.

Why was this blanket implementation adopted in such a short time frame? One cynical answer is the pressure to complete the roll out within the three year New Zealand election cycle. A second answer, given in discussion with ministry officials responsible for the implementation, is that there is no need to trial the policy and its implementation because it builds on what the system is already required to do. Unfortunately, the evidence shows that even if these requirements are understood by New Zealand teachers and leaders, there is enormous variability in the capacity to enact them, and that variability is likely to undermine the equity goals of the policy.

**Discussion**

School self management was introduced in New Zealand in 1989 because the government of the day believed that if school leaders had more freedom to run their school as they saw fit, and were more accountable to their local communities, schools would be more responsive to the needs of their students and performance would improve (New Zealand Department of Education, 1988). The accumulated evidence from over a decade of international surveys has

shown that this assumption was faulty. Our two examples have illustrated how the capacity in teaching and leadership that is required to resolve a longstanding problem in the NZ system – that of inequitable achievement outcomes for Māori, Pacific and low income students - is far greater than the capacity that is available in many of our self-managing schools. They have also illustrated how the structure and culture of school self-management complicates the systematic development of that capacity.

There is accumulating evidence about the characteristics of education systems which have sustained high and equitable performance or have made significant improvement in that performance. When six criteria derived from this evidence were used to evaluate two reform examples, it became clear that the NZ system is weak on one or more of them.

The recent introduction of national standards for the primary years does represent progress, in that ambitious achievement expectations for *all* students of a given age are now available to both teachers and parents. Assuming that the profession becomes more committed than it is at present to the value of those standards, they will provide a means for monitoring our collective progress towards national and school-level goals. This means greater potential to meet criteria such as professional forms of accountability and institutionalizing improvement. But there are problems in meeting other criteria, such as full system-wide commitment and developing sufficient capacity in schools and classrooms to meet the ministry's equity goals.

While the desired outcomes of New Zealand's reform efforts may be clearer, the examples have shown that the strategies required to achieve them are too weak, and too poorly specified to reliably achieve them. Although the Ministry required and supported clusters to engage in a SI process, they did insufficient to ensure that clusters made good choices about how to improve. Clusters were required and supported to develop a theory of improvement, but not necessarily a valid one; to select professional development providers, but not to select them giving due weight to coherence and demonstrated success; to collect data but not to share and use it in ways that build a strong professional community. Thus, while cluster activities were constrained by standardized contractual procedures, the constraints did not include the very qualities that would increase the probability of successful outcomes.

Why, in practice, was the requirement for effective, evidence-based decisions given so little weight in the SI example? Why did government and policymakers reject the trialing of the standards and roll them out without a detailed implementation plan? Definitive answers to

these questions would require a thorough policy analysis - which is well beyond the scope of this paper.

Our two illustrative examples do, however, suggest some possible explanations. Our argument is that systematic capacity building is in considerable tension with some key assumptions about school self management. Those assumptions are:

- That it is the responsibility of school leaders to manage their schools in ways that ensure student achievement at appropriate levels

- That central intervention to support this goal is only required in exceptional cases, because most schools have the capacity to achieve this goal

- That any central intervention must be delivered in a way that respects leaders' decision-making authority and avoids their dependence on outside help.

These assumptions create a culture in which any *externally-initiated* intervention is risky, because it inevitably communicates the message of exceptionality and failure. As correctly sensed by many SI coordinators and facilitators, the political and relationship risks of external intervention are high, regardless of the skill of the individual involved. Thus the state intervenes with one hand tied behind its back, and their cautious messages about what and who needs to improve contribute to lack of trust and defensiveness. School leaders may feel they have failed the system's expectation of self-managed success.

It is time to alter the default setting of school self-management – the view that the great majority of New Zealand schools can meet ambitious equity goals. A fairer and more realistic assumption is that the challenge of reducing New Zealand's inequitable distribution of achievement cannot be met by either single schools, or by clusters of schools that are coordinated rather than led, that have insufficient access to expertise, and insufficient oversight on issues of timeliness, quality and evidence of progress. Nor can this challenge be met by the provision of resources, no matter how smart they are, because resources do not change practice in deep ways unless they are mediated through face-to-face interaction (Halverson *et al*., 2004).

The default position should be that *all* schools will need help with this task and the exceptions will be those who can demonstrate that their Māori, Pasifika and low income students achieve at age appropriate levels on the national standards. By turning the default

position of adequate capacity on its head, the equity challenge will become a genuinely shared task in which teachers, students and the ministry can build trust and be mutually accountable (Bryk and Schneider, 2002; Elmore, 2004). While our two examples showed that there is a growing understanding among educators of the nature of their school's equity problem, the recurrent limitation of our interventions will persist until political and professional leadership communicate clearly about why the default position of sufficient capacity must be abandoned. When there is system-wide understanding of and commitment to that shift, ministry and schools will be able to work together to meet our equity goals with less ambivalence and defensiveness and with more urgency, focus and intensity.

# References

Annan, B. (2007), "A Theory of Schooling Improvement: Connectivity and Consistency to Improve Instructional Practice", Unpublished PhD thesis, University of Auckland, New Zealand.

Berliner, D. C. (2009), "Poverty and potential: Out of school factors and school success" available at: http://nepc.colorado.edu/files/PB-Berliner-NON-SCHOOL.pdf (accessed 5 July 2010).

Bryk, A. S. and Schneider, B. L. (2002), *Trust in Schools: A Core Resource for Improvement*, Russell Sage Foundation Publications, New York, NY.

Darling-Hammond, L. and Bransford, J. (Eds), (2005), *Preparing Teachers for a Changing World: What Teachers Should Learn and be Able To Do*, John Wiley, San Francisco, CA.

Elley, W. (2005), "On the remarkable stability of student achievement standards over time", *New Zealand Journal of Educational Studies*, Vol. 40 No. 1 and 2, pp. 3-23.

Elmore, R. F. (2004), *School Reform from the Inside Out: Policy, Practice, and Performance*, Harvard Education Press, Cambridge, MA.

Halverson, R., Kelley, C. and Kimball, S. (2004), "Implementing teacher evaluation systems: how principals make sense of complex artifacts to shape local instructional practice", in Hoy, W. K. and Miskel, C. G. (Eds), *Educational Administration, Policy and Reform: Research and Measurement*, Information Age Publishing, Inc., Greenwich, CT, pp. 153-188.

Koefoed, W. J. (2009), "Written Reporting: Strengthening Learning Partnerships through Purposeful Reporting", Unpublished PhD thesis, University of Auckland, New Zealand.

Kruse, S., Louis, K. S., and Bryk, A. (1994), "Building professional communities in schools", *Issues in Restructuring Schools Issue Report, No. 6*, Center on Organization and Restructuring of Schools, University of Wisconsin-Madison.

Lai, M. K., McNaughton, S., Amituanai-Toloa, M., Turner, R. and Hsiao, S. (2009), "Sustained acceleration of achievement in reading comprehension: the New Zealand experience", *Reading Research Quarterly,* Vol. 44 No. 1, pp. 30-56.

Lai, M., McNaughton, S. and Hsaio, S. (2010), *Building Evaluative Capability in Schooling Improvement*, *Milestone Report, Part B: Strand Two*, Ministry of Education, Wellington, New Zealand.

Le Fevre, D., (2010), "Changing tack: talking about change knowledge for professional learning", in Timperley, H. and Parr, J. (Eds), *Weaving Evidence, Inquiry and Standards to Build Better Schools*, New Zealand Council for Educational Research, Wellington, New Zealand, pp. 71-92.

Levin, B. (2008), *How to Change 5000 Schools: A Practical and Positive Approach for Leading Change at Every Level*, Harvard Education Press, Cambridge, MA.

Levin, B., and Fullan, M. (2008), "Learning about system renewal", *Educational Management Administration and Leadership*, Vol. 36 No 2, pp. 289-303.

McNaughton, S. (2002), *Meeting of Minds*, Learning Media Limited, Wellington, New Zealand.

McNaughton, S. and Lai, M. K. (2010), "The learning schools model of school change to raise achievement in reading comprehension for culturally and linguistically diverse students in New Zealand", in Johnston, P. H. (Ed.), *RTI in Literacy – Responsive and Comprehensive*, International Reading Association, Newark, DE, pp. 313-336.

Mourshed, M., Chijioke, C. and Barber, M. (2010), *How the World's Most Improved School Systems Keep Getting Better*, McKinsey & Company, London.

National Organisation (2008), "*Policy: Education*", available at: http://www.national.org.nz (accessed 16 July 2008).

New Zealand Department of Education (1988), *Tomorrow's schools: The reform of education administration in New Zealand*, Government Printing Office, Wellington, New Zealand.

New Zealand Education Review Office (1996), *Improving Schooling in Mangere and Otara*, New Zealand Government, Wellington, New Zealand.

New Zealand Government (1989), *Education Act (Public Act 1989, No. 80)*, Government Printing Office, Wellington, New Zealand.

New Zealand Government (2001), *Education Standards Act (Public Act 2001, No. 88)*, Government Printing Office, Wellington, New Zealand.

New Zealand Ministry of Education (2007), *The Literacy Learning Progressions*, Learning Media Limited, Wellington, New Zealand.

New Zealand Ministry of Education (2008), "Ka Hikitia: Managing for success", available at: http://www.minedu.govt.nz/theMinistry/PolicyAndStrategy/KaHikitia.aspx (accessed 14 February 2010).

New Zealand Ministry of Education (2009a), *Reading and Writing Standards for Years 1-8,* Learning Media Limited, Wellington New Zealand.

New Zealand Ministry of Education (2009b), "Ministry of Education National Standards: Questions and Answers", available at: http://www.minedu.govt.nz/theMinistry/EducationInitiatives/NationalStandards.aspx (accessed 19 July 2010).

New Zealand Ministry of Education (2009c), "New Zealand Education Gazette", Ministry of Education, Wellington, New Zealand.

New Zealand Ministry of Education (2010), "PISA 2009: Our 21st century learners at age 15", available at: http://www.educationcounts.govt.nz/publications/series/2543/pisa-2009/pisa-2009-our-21st-century-learners-at-age-15/ (accessed 20 December 2010).

Odden, A. R. (2009), *10 Strategies for Doubling Student Performance*, Corwin, Thousand Oaks, CA.

OECD (2010a), "PISA 2009 results: Overcoming social background – Equity in learning opportunities and outcomes (Volume II)", available at: http://dx.doi.org/10.1787/9789264091504-en (accessed 14 February 2010).

OECD (2010b ), "PISA 2009 results: What students know and can do (Volume I)", available at: http://dx.doi.org/10.1787/9789264091450-en (accessed 14 February 2010).

OECD (2010c), "Strong performers and successful reformers in education: Lessons from PISA for the United States", available at: http://www.edweek.org/media/gps-us-strong-performers-and-successful-reformers.pdf (accessed 14 February 2010).

Parr, J. and Timperley, H. (2010), "Feedback to writing, assessment for teaching and learning and student progress", *Assessing Writing*, Vol. 15 No. 2, pp. 68-85.

Parr, J., Le Fevre, D., Lai, M., McNaughton, S., Hsiao, S., Timperley, H., Annan, B., Fisher, A., Kofoed, W. and Mose, K. (2010), *Building Evaluative Capability in Schooling Improvement: Milestone Report 8*, Ministry of Education, Wellington, New Zealand.

Raphael, T. E., Au, K. H. and Goldman, S. R. (2009), "Whole school instructional improvement through the standards-based change process", in Hoffman, J. and Goodman, Y. (Eds), *Changing Literacies for Changing Times*, Routledge, London, pp. 198-229.

Robinson, V. M. J., Lloyd, C. and Rowe, K. J. (2008), "The impact of leadership on student outcomes: an analysis of the differential effects of leadership type", *Educational Administration Quarterly*, Vol. 44 No. 5, pp. 635-674.

Robinson, V. M. J., Ward, L. and Timperley, H. (2003), "The difficulties of school governance: a layperson's job?", *Educational Management & Administration*, Vol. 31, pp. 263-281.

Sleegers, P., Geijsel, F. and Van Den Berg, R. (2002), "Conditions fostering educational change", in Leithwood, K. and Hallinger, P. (Eds.), *Second International Handbook of Educational Leadership and Administration*, Dordrecht, Kluwer, Vol. 1, pp. 75-102.

Timperley, H. and Parr, J. (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.

Timperley, H., Hohepa, M., Keegan, P., Parr, J., Lai, M. and McNaughton, S. (2010), *Building Evaluative Capability in Schooling Improvement: Milestone Report, Part A: Strand One and Strand Three*, Ministry of Education, Wellington, New Zealand.

Timperley, H., Parr, J., Hohepa, M., Le Fevre, D., Lai, M., Dingle, R. and Schagen, S. (2008), *Findings from the Inventory Phase of the Building Evaluative Capability in Schooling Improvement Project*, Ministry of Education, Wellington, New Zealand.

Wilkinson, R. G and Pickett, K. (2009), *The Spirit Level: Why More Equal Societies Almost Always Do Better*, Allen Lane, London.

Wylie, C., Hogden, E. and Barr, C. (2009), *National Standards Consultation and Analysis: Report for the Ministry of Education*, Council for Educational Research, Wellington: New Zealand.