A model of school change for culturally and linguistically diverse students in New Zealand: a summary and evidence from systematic replication

Stuart McNaughton and Mei Kuin Lai

Woolf Fisher Research Centre

The University of Auckland

June 2008

The Local Context

New Zealand has a deservedly high international reputation for literacy teaching and the levels of achievement of its students are high in international comparisons such as PISA and PIRLS. But New Zealand also faces a significant challenge in its literacy teaching which has two features. There is a long 'tail' in the distribution of achievement. Differences between the lowest quartile and highest quartile of students are very large in international terms. Secondly, two groups of students are over represented in this tail; they are Māori students (from the indigenous community) and Pasifika students (from recent immigrant, and second and third generation Pacific Island families). These students are most often found in schools serving the communities with the lowest incomes levels and lowest employment levels.

New Zealand researchers and policy makers have known for some time that we have not served these communities as well as others and that the schools are less effective with these communities. With exceptions in some schools, the period following colonisation has been associated with schooling that has not been equitable for indigenous students. With waves of immigration of Pasifika families from the 1960s, a new set of communities for whom teaching is less than effective have been identified. The schools of South Auckland which have high proportions of Māori and Pasifika students from communities with the lowest employment and income levels have been identified in policy terms as sites of low achievement and in crisis. More than twenty years ago the title of a report by Ramsay, Sneddon, Grenfell & Ford (1981) proclaimed the crisis was such that "Tomorrow may be too late". The international comparisons through the 1990s added to the starkness and seeming intractability of this failure to deliver.

Starting in 1998, a policy context and associated resources and intensive research and practice endeavors have contributed to a marked reduction in the disparities in levels of achievement in early reading between Māori and Pasifika students and other students. This change has been identified in national monitoring at year 4 of schools in cycles of assessments from 1996 to 2000 to 2004, with the latter two years showing percentages of year 4 children reading below their age levels dropping from 19% to 12% and then to 7% over those cycles. Despite this, the evidence also suggests that at Year 4 and Year 9, the disparities in reading comprehension have continued, if not increased (Crooks & Flockton, 2005).

In a series of studies, researchers from the Woolf Fisher Research Centre in partnerships with schools, their communities and policy personnel have focused research and development programmes on this pressing and long standing educational challenge for reading comprehension. The major assumptions behind this work has been that instructional effectiveness in the schools could be increased so that achievement levels in reading comprehension for Māori and Pasifika students in school years 4 through 8 were substantially accelerated; that the change would require a model of schooling improvement that solved the challenges of effectiveness in context through building the expertise of the teachers; and that educationally significant and sustainable changes needed long term partnership between researchers, policy makers and school professionals.

The criteria we have set for educationally significant changes include accelerated rates of achievement (changing levels but not changing rates can mean groups of students never 'catch up'), and shifting the distribution of achievement so that the achievement in the schools matches the national distribution (the probability of being in any one part of the distribution such as high or low or average bands is no different for these students than what would be expected nationally). There are several meanings for the term 'sustainable' but in the context of these targeted clusters of schools it means maintaining rates of change in achievement as well as maintaining effective problem solving of local challenges to being effective for Māori and Pasifika students.

The International Context of 'Schooling Improvement'

The evidence base for schooling improvement does not suggest we could be very optimistic about substantial and long lasting changes. New Zealand's and other countries' response to the enduring "education debt" of less than effective schooling for 'minority' children has included programmes of schooling improvement and school reform at local, district, and even national levels. There is evidence for varying degrees of effectiveness for these programs.

In the United States by early 2000 over 100 different comprehensive school reform models aimed at improving instruction and student achievement had been developed and implemented in between 10-20% of the elementary schools (Rowan, Camburn, Correnti & Miller, 2007). Less than a third of these (Borman, 2005) have been widely replicated sufficiently to enable systematic and reliable analyses of effects. Just like the earlier large scale attempts in Head Start to change

the educational lives of poor and minority children in the 1960s which involved many models, this massive investment in educational change has produced mixed models with mixed results.

Borman (2005) reviewed the evidence for scaled-up projects of school reform in high-poverty schools in the United States. The review shows that they produce widespread, but generally modest effects (effect sizes between 0.1 and 0.2). Although initial effects are limited, the evidence also suggests stronger effects appearing after 5 years. In addition, researchers at the University of Michigan have been comparing three types of the most widely disseminated and sustained programmes, varying in terms of specificity (how detailed the specifications are for the whole programme) and scriptedness for teachers (the degree to which lessons are prespecified and detailed in terms of exact moves and sequences for teaching), and comparing them with each other and with schools with no programmes. They have examined implementations in over 100 high poverty schools with over 7,000 students.

The two programmes that are highly specified and targeted on specific domains in reading and writing are more effective than no programmes and the programme that works at a generic level around ideas of teaching and learning. But there are differences between the two more specified programmes. One is a highly scripted and externally designed programme which is focused on explicit skills instruction and the other which also develops explicit instructional modes is designed collaboratively with external advisors and professional leaders within the school professional community and has a more literature based focus on reading and writing across genres and higher order comprehension. The former highly scripted programme was very effective for beginning reading instruction but not for later achievement in reading and writing at grade 5. The second programme was very effective at the later stages of reading and writing at school.

From these and the analyses such as Borman's (2005), generalisable principles of effectiveness are being derived; for example about the need for programme specificity, the role in consolidating change of professional learning communities in schools; how the level and quality of implementation affects changes; the need for close and reciprocal relationships between the developer and the local school and school district; and the need for the coordination and fit of the model to local circumstances. But the evidence for large scale generalisable changes against the criteria we noted before (page 3) is very limited.

The Model of Change and Evidence

We have tested a model of changing schools to be more effective. The model is based on several key principles. These are reviewed below together with a description of how these components are operationalised. The evidence for the effectiveness of this model is summarised in the final section. The research and development programme has been systematically implemented and then replicated across 3 clusters of schools involving some 7000 students and 200 teachers in any given year. It has focused almost exclusively on schools and teachers' instruction. We assume that to optimise school effectiveness there would be components that are focused on the home and school relationships and that the linkages between communities and schools would contribute a substantial amount to the effectiveness of the schools. Our programme of research has deliberately not focused on this part of the ideal school adopting the research decision to better understand the components within the schools. Members of our research team continue to examine community literacy and language practices and relationship with schools. But this strategic decision for the school change research has been an attempt to gain some evidential control over what the school community and instructional side of that relationship is able to accomplish under the best model that we can design through a research and development process.

Key Principles of the Model of Change

This model of change is drawn from the theoretical and research literature in several areas. These are about the nature of teachers and teaching, of schools and their professional communities, the nature of programmes of instruction for literacy, the nature of research and development partnerships and the nature of sustainability and scalability. From these we have extracted several key principles.

1. Teachers need to be able to act as adaptive experts.

Recent policy work in New Zealand (and in other countries), based on analyses of the variance in achievement attributable to such elements as student background, the nature and types of schools and the classrooms conclude that the key to school change is more effective teaching (Alton-Lee, 2004). More effective teaching requires teachers to act in particular ways. In our model teachers are seen and treated as professional experts and teaching is a form of expertise. Experts are deeply knowledgeable about what they do, how they do it, and why they do it. Their knowledge and skills are about particular practices of literacy using particular forms of guidance and assessment tools in particular settings. Like expertise generally, this has three components: a knowledge base (articulated knowledge and understanding of the domain of teaching and of

children and their learning and how to teach effectively), strategic practices which have versatility and adaptability; and a metacognitive component involving keen awareness of the effectiveness of one's practices through both reflection and regulation of these practices.

General models of expertise identify how experts are goal focused and intentional; they are strategic, being able to adapt to circumstances; and they are keenly aware of the effectiveness of their performances in the sense of being in control by being able to monitor, check and modify (McNaughton, 2002). They do this on the basis of extensive and well articulated knowledge of the domain within which they are expert. So too teachers need extensive knowledge about the domain (the content area such as literacy and language), how and what to teach (pedagogical content knowledge) and knowledge related to students and communities. In the context of linguistically and culturally diverse students the latter entails an understanding of their children's language and literacy practices as these reflect children's local and global cultural identities. Importantly this means knowing how these practices relate (or do not relate) to classroom practices (New London Group, 1996).

These general attributes give experts the twin features of being technically adept as well as innovative and adaptable, although Darling-Hammond and Bransford (2005) differentiate between the relative weighting of these attributes in their descriptions of new forms of teaching for a changing world. They see teachers as acting along a continuum from routine experts who accurately employ scripted lessons through to adaptive experts who selectively and strategically apply known instructional procedures but are constantly refining and changing to be more effective. The former develop a core set of competencies that they apply with greater and greater efficiency, while the latter continuously add to their knowledge and skills. These latter experts are innovators; they change core competencies and expand the breadth and depth of their expertise. The description of adaptive experts is more like the general view of expertise outlined above and our position is consistent with theirs in claiming that for teachers to be more effective with culturally and linguistically diverse students, teachers need to be more like adaptive experts.

This is a different view from those schooling improvement programmes which design and implement highly scripted lessons which teachers then efficiently apply. This programmatic approach is based on the need to guarantee treatment fidelity (that successive implementations do in fact conform to the programme design) and scalability (rolling out the same programme with known attributes and predictable effects). As noted above, the evidence is that programmes with

such routinised forms of teaching (such as those in *Success for All*[1]) can guarantee high levels of beginning reading skills for minority children in poor schools. But highly scripted programmes may be less effective at later stages of reading and writing. The reasons are that using scripted routines for what an eminent researcher in literacy, Paris (2005), calls 'constrained' skills means that these generally can be taught and learned quickly and efficiently because they are a closed set (there is finite set of items to learn such as letters or sounds). But the more open 'unconstrained' skills which are the basis for reading comprehension and writing for different purposes, that is the more language dependent skills, require much more adaptable and complex forms of teaching. These features of teaching are undermined with closely scripted approaches. Also, becoming skilled in the beginning stages is a necessary but not sufficient condition for later complex thinking and learning such as comprehending and critically understanding texts.

2. *Local evidence about teaching and learning is necessary to inform instructional design.*
Unlike some other approaches the model assumes the need to examine local 'evidence' in order to design effective instruction. Using localised or contextualised evidence about both teaching and learning is needed at two levels. One is a requirement to base instructional practices collectively on evidence about teaching and learning drawn from on-the-ground patterns. The second is the requirement that an individual teacher is able to use a range of assessments both formal and informal and a range of contexts both formal and informal to broaden his or her knowledge of individual students and effectiveness with those students in order to better personalise instruction.

Paradoxically, the need to localise the 'evidence' at both these levels comes because we have a well developed research base about the nature of reading comprehension and teaching. Generally, there is considerable consensus around what students need to learn and what effective teaching for reading comprehension looks like. In order to comprehend written text a reader needs to be able to decode accurately and fluently, to have a wide and appropriate vocabulary, have appropriate and expanding topic and world knowledge, active comprehension strategies, and active monitoring and fix up strategies (Block & Pressley, 2002). So it follows that children who are relatively low progress may have difficulties in one or more of these areas. The consensus around effective teaching identifies attributes of both content (curriculum) and process (Taylor,

---

[1] Success For All uses a prescriptive approach to school change increasingly exponentially implemented school by school. The first step occurred when in 1986. The design of the reading programme as it occurs in the classroom is specified, including the sequencing and teacher actions within the sequence. It is described as highly scripted. (Slavin & Madden, 2001)

Pearson, Peterson & Rodriguez, 2005). In the middle grades these attributes include instructional processes in which goals are made clear, and which involve both coaching and inquiry styles that engage students in higher level thinking skills. Effective instruction also provides direct and explicit instruction for skills and strategies for comprehension. Effective teaching actively engages students in a great deal of actual reading and writing and instructs in ways that enable expertise to be generalisable and through which students come to be able to self regulate independently.

In addition, researchers have also identified the teacher's role in incorporating cultural resources including event knowledge (McNaughton, 2002) and in building students' sense of self efficacy and more generally motivation (Guthrie & Wigfield, 2000). Quantitative and qualitative aspects of teaching convey expectations about students' ability which affect their levels of engagement and sense of being in control. Culturally and linguistically diverse students seem to be especially likely to encounter teaching which conveys low expectations (Dyson, 1999). There are a number of studies not directly of reading comprehension but in schooling improvement which have shown how these can be changed and how they impact on instruction and learning. In general, both changes to beliefs about students and more evidence based decisions about instruction are implicated, often in the context of school wide or even cluster wide initiatives (Phillips, McNaughton & MacDonald, 2004).

It follows that low progress could be associated with a variety of teaching and learning needs in one or more of these areas. Out of this array of teaching and learning needs, those for students and teachers in any particular instructional context may therefore have a context specific profile. While our research-based knowledge means there are well established relationships, the patterns of these relationships in specific contexts may vary. A simple example might be whether the groups of students who make relatively low progress in a particular context, say a cluster of similar schools serving similar communities, have difficulties associated with decoding or with use of comprehension strategies or both, and how the teaching that occurs in those schools is related to those difficulties. Buly and Valencia (2002) provide a case study from a policy perspective of the importance of basing any intervention on specific profiles, rather than making assumptions about what children need (and what instruction should look like). In that study, mandating phonics instruction for all students in the State of Washington who fell below literacy proficiency levels had missed the needs of the majority of students, whose decoding was strong but who struggled with comprehension or language requirements for the tests. Using information

about both student learning and achievement and about teaching and instruction is needed to avoid making mistakes including those described by Buly and Valencia (2002).

3.   School professional learning communities as vehicles for changing teaching practice

Our model locates teachers and teaching in a professional school community which at the beginning of a process of change not only includes teachers, leaders and other school professionals, but also researchers and local district managers (in New Zealand's case national Ministry of Education personnel). Each of these members learns to act and practice effectively within those communities.

The significance of professional learning communities to improve teaching practices and student achievement has been recognised for some time and specific attributes are linked to improvements in teaching, student achievement and student learning (e.g., Louis, Marks & Kruse 1996; Robinson & Lai, 2006). Approaches in which communities focus on collective problem-solving around agreed evidence result in sustainable improvements in student achievement particularly, reading comprehension. For example, when Cawelti and Protheroe (2001) examined the factors responsible for student achievement gains in six formerly underperforming districts in the United States with successful school improvement efforts, one of the attributes of a successful district was teacher analysis and use of achievement data. Over eight years, district-wide pass rates in the state reading, writing and mathematics tests rose from a baseline of between 65% and 70% to 95% with reductions in disparities among ethnic minorities in low socio-economic groups. Similarly, a follow up of a New Zealand schooling improvement intervention showed that schools that maintained the substantial achievement gains and/or built on them after the intervention emphasised raising achievement and the collegial analysis and discussion of student achievement data to change teacher practice (Timperley, Phillips & Wiseman, 2003). Apart from these emphases, there was no apparent relationship between a school's achievement level and other variables under investigation such as class size, teacher turnover, the value the school placed on professional development and the level of implementation.

Improvements in achievement are not associated with sharing ideas but with critical reflection on practice. The features of professional learning communities that can effectively analyze evidence to improve teaching practices and raise student achievement have been identified (Coburn, 2003; Robinson & Lai, 2006). One is collective inquiry to improve teaching and learning (Seashore-Louis, 2006; Timperley, Wilson, Barrar & Fung, 2008). Inquiry requires teachers to examine

current student learning, reflect on the teaching and learning that needs to occur to improve current student learning, develop appropriate practices to address the identified needs and monitor student progress. To do this requires the deep knowledge associated with teacher expertise noted earlier (pp. 5-7). The theoretically rich inquiry requires consideration of theories researchers might bring to the professional community as well as the engagement of teachers' tacit and explicit ones (Robinson & Lai, 2006; Timperley et al., 2008). Engaging the teachers' theories means uncovering the reasons and conditions that have resulted in their current practices and this process is strongly linked to interventions which have improved achievement (Ibid, 2008).

The process of effective inquiry requires participants in the professional learning community to learn to adjudicate between differing theories for the patterns in achievement by carefully examining profiles of students needs to test their theories. In one study, the professional learning community adjudicated between two possible causes for low achievement in reading comprehension by examining profiles of students needs using standardised tests. The profiles indicated that students were high decoders but weak in other aspects of reading comprehension thereby discounting one theory (that students needed more support in decoding) and confirming the other (that students' decoding was adequate but they needed support in comprehending what they were decoding) (Robinson & Lai, 2006). This emphasises again the need for close descriptions of student learning as well as descriptions of patterns of teaching, which are sufficiently broadly based to provide a basis for informed decisions about teaching, collected and analyzed together within the community to clarify and test hypotheses about how to develop effective and sustainable practices (Phillips, McNaughton & MacDonald, 2004).

An analytic stance to the collection and use of evidence is required in that members need to know whether and how planned interventions do impact on teaching and learning and how these interventions are meeting their goals. The research framework adopted by the community needs therefore to be staged so that the effect of interventions can be determined. The design part of this is by no means simple especially when considered in the context of recent debates about what counts as appropriate research evidence (McCall & Green, 2004). Our model assumes that research that makes a difference to schools requires designs that are appropriate to the 'messy' conditions of schools in which change is a constant (in such things as the members as well as the students and in curricula and resources), and in which the communities' members hold values, beliefs and ideas and who engage in practices that reflect these. We have developed a robust

framework which has both quantitative and qualitative features and employs a robust quasi experimental design.

Whilst it is possible for individuals or individual schools to engage in inquiry, having wider collective inquiry involving schools, researchers, and policy personnel means that different forms and degrees of expertise is brought to these communities. The distribution of their respective expertise means that as members contribute to collaborative practices they fit their knowledge, strategies and reflection to the effectiveness of those practices and in so doing define, add value to and extend the practices (Wells, 1999). Researchers may bring aspects of analytic skills and theoretical understanding to the community. Teacher leaders may bring understanding of how assessment practices and professional development are linked and teachers may bring deep knowledge of how specific instructional practices work and can be adapted within their classrooms. In addition, the variation in expertise in the community's practices due to newness in the community or limited experience or knowledge can be supported by the collective enabling the novice participants to more from being peripheral to the community to more central and knowledgeable members (Lave, 1991). Teachers may initially need to be supported though being more like routine experts but as they come to be knowledgeable and through the guidance of and collective engagement with other members they can develop as more adaptive experts. In that sense the community is a collective to which new members, including new leaders and new researchers are apprenticed. In Darling-Hammond and Bransford's (2005) model of teacher expertise, understanding of how communities function in their norms and modes of operating is a core component of what teachers need to learn. This applies not just to creating a community of learners in a classroom but coming to know how to contribute to a professional community.

Collective inquiry is also important to developing both a collective as well as person sense of efficacy (Bandura, 1995). (Self-efficacy is the belief in one's own ability to produce the desired outcome and collective efficacy is the collective's belief in its ability to produce the desired outcome). Where teachers' talk is mostly on the seemingly insurmountable or intractable difficulties and where that talk locates the locus of the difficulties in areas outside of the teachers' immediate control such as employment, low parental interest, housing or language 'difficulties' (all of which in some sense may need to be better understood for our schools to be more effective, but may be outside of a teacher's immediate control) there is a low collective sense of efficacy that undermines teacher's personal sense of efficacy. Bandura (1995) describes a depressing cycle in which a sense of unsolvable problems lowers beliefs in personal effectiveness, which in turn

results in a decreased commitment to teaching and innovating. The resulting impact on students further reduces teachers' sense of being effective. At the other end of a continuum is there is a high sense of collective efficacy where teachers collectively share a sense of solving difficulties and of making a difference, and have the evidence to support these beliefs. High levels of collective efficacy at the beginning of an academic year predict a school's academic achievement at the end of the year. This can be demonstrated even when different characteristics of children, their prior levels of achievement and staffs' levels of experience are controlled. A strong sense of personal efficacy is also associated with effective teaching with culturally and linguistically diverse students (Bandura, 1995).

There are several major influences of the development of teachers' personal sense of efficacy. These include experience of mastery as well as immersion in an effective community. The most powerful influence is the experience of having being effective in the past.

4. *Educative research – practice - policy partnerships*

The features of school-based professional learning communities noted above require educative partnerships between researchers, practitioners and policy makers that are focused on collectively learning how to improve teaching and learning. As noted in the previous section, this allows for multiple sources of knowledge and expertise from the research and policy communities to be brought to a learning community, thereby increasing the possible pool of expertise to address the teaching and learning needs (Annan, 2007). It is possible for partners to work together in ways that are not educative. For example, policy-makers might create policies that are not aligned to the emerging research on effective teaching or researchers might not understand a programme they have designed cannot be effectively implemented as designed because the programme rely on high teacher stability in the school. So learning for all parties can be sped up if they each use their respective expertise to collectively solve the pressing educational issues.

In an early evaluation of a schooling improvement initiative, Timperley, Robinson & Bullard (1999) found that partnerships between local communities, schools and government were highly problematic for reasons such as blaming another partner for the educational "failures", rather than attempting to learn together how best to raise achievement. We have avoided this in our work by using the approach outlined in Robinson and Lai (2006), where each party is able to mutually critique each other's emerging tacit and formal theories of how to address students learning needs. Using this approach, no one theory is privileged over the others allowing each partner

equal power in decision-making. For example, researchers supported schools uncover the need to check for evidence, but at the same time, ensured that their theory of checking was open for critique (Ibid, 2006).

This raises two important considerations. Firstly that the policy environment must be responsive enough to learn from other partners. Annan's (2007) analysis of the policy environment of the United States of America, United Kingdom and New Zealand suggests that hierarchical 'top-down' policy structures like the USA are less open to learning from other parts of the hierarchy. A 'flat' policy structure like New Zealand allows for multiple opportunities for engagement with national and local policy-makers.

A second consideration, is the engagement of the wider parent and family community as a major partner. In the school clusters we have worked with there is considerable community involvement as the governance of New Zealand schools is via local communities (Annan, 2007). However, in the recent research and development sequence we have not directly focused on this part of the partnership. In the next stages of our research we are planning more direct research and development work with communities for example around how schools can contribute to and learn from community practices which are related to gains during school years and over summer (between school years).

5. *Instructional leadership*

The importance of teacher leadership in raising achievement has been highlighted in recent studies (Robinson, Lloyd & Hohepa, 2007). However it is not leadership per se that will support the improvement of achievement, rather particular dimensions of leadership that lead to greater impact on student outcomes. Robinson et al.'s. (2007) review of the impact of leadership in student outcomes highlights five leadership dimensions associated with higher achievement - establishing goals and expectations including the setting, communicating and monitoring of learning goals; strategic resourcing which aligns resource selection and allocation to priority teaching goals; planning, coordinating and evaluating teaching and the curriculum; promoting and participating in teacher learning and development; and ensuring an orderly and supportive environment.

Our model is based around lead teachers, who are responsible for leading the interventions in the school, and relies on them working effectively with the Principal and other senior managers to

implement the intervention in the school. Whilst principals are not directly involved in leading the intervention, they have considerable influence on the shape of the intervention in the school, as our interventions are designed to work through the existing school system, albeit to change those in partnership with school leaders if necessary. Unlike interventions that are highly prescribed and may be less open to variability in leader capacity, the interventions that we have implemented are influenced by how teacher leaders understand and enact their roles.

The function of effective leadership can be illustrated be three examples from our studies. First, schools had to resource about half of the intervention themselves so leaders had to strategically plan their budgets to prioritise funding the interventions, particularly the part of our intervention where every teacher was involved in fortnightly workshops. This also meant careful timetabling to ensure classes were covered during this time, which the schools all managed to coordinate. Second, each lead teacher collaborated with the researchers and each other to plan, coordinate and evaluate the effectiveness of the intervention in their schools through the continual monitoring of the impact of their practice on student achievement. In one cluster of schools, the teacher leaders also developed a cluster-wide observation plan to observe teachers implement the intervention. Finally, all lead teachers were required to participate in every part of the professional development in order to be able to implement it in their schools. Principals were also involved, albeit indirectly through the cluster-wide principals' clusters and other such mechanisms.

In our model the professional community at the level of a school or a cluster of schools has a further responsibility to contribute to the development of these attributes of the leaders. We cannot assume that teacher leaders will automatically have such attributes and knowledge, for example we have found mixed capabilities to self-review using evidence (e.g., Timperley et al., 1999). Similarly, the model is vulnerable to changes in leadership unless it enables communities to build effective induction systems for new members.

6. *Existing approaches can be fine tuned, or modified to design more effective forms and functions of instruction for Māori and Pasifika students.*

Our model is built on a general proposition that rather than importing a new programme for reading comprehension instruction, as it were getting a programme off the shelf, the existing programme in schools could be modified to be more effective. Partly this comes from evidence about the practices that are already in place, and partly from theoretical decisions about the nature of changing teachers' instruction given the view of teachers as adaptive experts. More

specifically, one reason is that in international terms New Zealand's literacy instruction is relatively effective, seen in professional international opinion as well as in the international studies with such measures as the proportion of students who are in the highest achievement bands and overall performance in different areas of literacy (e.g., PISA and PIRLS). There are known features of current practices that are able to be used effectively with culturally and linguistically diverse students. These include the texts and the language base to those texts (McNaughton, 2002). That means that the existing knowledge and practices can be used as a lever for more effective practices.

A second reason is that in analysing the evidence in our programme we have been able to show that considerable variability in effectiveness exists across teachers and schools. There are some teachers and schools who are relatively more or less effective. It is in the variability that exists that one can identify and use as exemplars new knowledge and practices from other teachers, again leveraging off existing practices. A third reason derives from a basic idea in effective change. Effective education starts from the repertoire that exists and builds greater expertise from that. This is basic principle in professional development as much as in beginning literacy instruction. The fourth reason is that designing a whole scale change in programme unless it is consistently and impressively implemented runs the risk of being counter productive to the need to treat, and move teachers towards being, adaptive experts.

7. *High treatment integrity or treatment scalability and sustainability are possible*

Research in schooling improvement is currently grappling with the issue of sustainability and the closely linked ideas of treatment integrity or fidelity and scaling up. Treatment fidelity and integrity refer to the degree to which in any intervention or development, a defined treatment is actually applied as designed. The associated concept of scaling up is taking a defined program of change or treatment and applying it across new contexts, again assuming known features of a programme with known results are successfully put in place. While sustainability has several meanings, the most usual is akin to generalization across time (the effects of the intervention keep going to some defined level). But sustainability can also mean sustaining the treatment with new cohorts of learners and also with new cohorts of teachers. The latter meanings are closer to the idea of integrity and fidelity.

There is an issue here for schooling improvement. Integrity has been associated with the degree to which the programme is specified for a school or sets of schools and this specificity has been seen as very significant for two reasons. One is because changing teachers' practices requires clarity and elaborateness on the part of the design team (Cohen & Ball, 2007), and the other is the need to guarantee effects, the core requirement of treatment integrity (Coburn, 2003).

In schooling improvement the aim has been to scale up through guaranteeing high fidelity, and consequently bemoaning adaptation. But adaptation is an inherent property of teachers' responses to new ideas. Datnow and Springfield (2000) note that even with externally developed reform designs, a process of fitting the design to local circumstance takes place at both school and district level. Lefstein (2007) argues that the partial implementation of the National Literacy Strategy in England was determined by the pedagogical beliefs of the teachers. Teachers cobbled new ideas onto existing practices that reformers were trying to supplant.

This sort of mixed implementation is quite generally seen as problematic and is not limited to the schooling improvement literature. Davis and Sumara (2003) voice this typical frustration in teacher professional development. In their study teachers used the vocabulary of constructivism to plan, justify and reflect on their teaching. But what they did often bore little relationship to the core original constructivist frameworks. Partly this is because in interventions ideas are not adequately articulated (Cohen & Ball, 2007), but it is also because of the ways teachers reconstruct and reframe their practices in idiosyncratic ways.

So, our model which fits an intervention to local circumstances seems to be in tension with these needs to be very specific. The model reflects a tension associated with teacher learning within schooling improvement (Coburn, 2003). It is between importing a set of procedures in a way that risks undermining local autonomy and efficacy and a more collaborative development of common procedures which risks losing instructional specificity.

The research literature on schooling improvement has become ambivalent about these issues. Reviews point out on the one hand that a high degree of program prescription (specificity) is important (Borman, 2005). Reviewers also argue that approaches in which professional development focuses on joint problem solving around agreed evidence, such as student achievement outcomes, are more likely to result in sustainable improvements in student achievement.

We solve these tensions in the following way. Firstly, we distinguish between the degrees of prescription or predetermined specificity of the content of what to teach and the specificity of the process of change. The important point is that the process of change (through features such as inquiry and analysis of evidence), can still lead to a programme that can be very specifically described in what is taught and how it is taught. This specificity in the content comes from a relatively more open-ended starting point. The distinction here means we can distinguish between two sorts of fidelity too. One is fidelity of a program and another is fidelity of a process which can still deliver a high degree of program specificity which has been localised. Our objective in the model is to then to build fidelity to the content or programme across classrooms and schools based on the process.

8. *Components of the process of change need to be introduced and developed over time.*

The change model is operationalised in a three phase design in which the phases add components within an applied research (quasi experimental) design. The first phase involves the collection and analysis of student learning including achievement and progress data. It also involves the collection and analysis of classroom observational data. Together these provide profiles of teaching and learning from which a number of activities are possible. The 'evidence' can be used to obtain a closer analysis of the strengths and weaknesses of learning and development of literacy for students at various levels, for example individual classrooms, year levels, ethnic groups, schools. Current instructional approaches can be matched to the students' profiles to make judgments of where the teaching appears to be less than effective and where there are strengths. In each case sources of positive deviance or outlier classrooms, teachers and perhaps schools where the patterns of low achievement are not present can be identified for further shared analysis. This phase also builds up the communities' shared beliefs and expectations about the role of teachers, researchers, leaders, polity makers and the shared practices around enquiring into and sharing and using evidence. The phase also is important in research design terms to give a baseline against which changes can be compared.

The second phase continues the evidence collection but adds targeted professional development using the profiles as a means to design the particular focus of the teaching programme. The third phase continues the evidence collection and use and shifts from the professional development mode to one of sustaining the processes of instructional change and further problem solving. For

example, we collectively design systems for teacher groups to plan topics or teaching sequences together and observe and reflect on each other's implementations. Inquiry projects are developed and the results reported at teacher led cluster conferences.

It could be argued that if the needs for teaching and learning are clear (for example known patterns in very similar schools) then there may be efficiencies to be made in reducing phases and concentrating on the professional development part. However, the need to have a well structured process that builds capacity in the senses noted above and also the capability to carry out and maintain more effective teaching, even if replicating a well designed programme, is signaled in the research literature. The most telling researched example comes from the attempt replicate the programme of reform carried out in the urban schools of New York City's Community School District #2. The District #2 reform had been highly successful. Under the leadership of Superintendent Anthony Alvarado student achievement ranking rose from near the bottom of the thirty-two community school districts in 1987 to second.

In 1998 the San Diego City Schools attempted to reform their schools using the New York model. The goal of the San Diego reform was the same, and the architect of the initial reform in New York, Superintendent Anthony Alvarado, was brought in to oversee it. There were some similarities between the districts and some notable differences, but what is clear is that by 2002 the replication had not delivered what was expected and it needed a major overhaul and change of direction. It was apparent that the scaling up had limitations. Hubbard, Mehan and Stein (2006) painstakingly analysed this at scaling up at each of the intersecting levels of the district policy and organisation, schools and classrooms and teacher student interactions. The picture that emerges is that it was flawed by the whole sale adoption of the reform as a template without adopting the reform process as a developmental process.

Summary of Research Evidence

Our first research and development project focused on reading comprehension, conducted as a collaborative partnership between researchers, schools and the New Zealand Ministry of Education, was designed to develop and test the model in a cluster of 'decile 1' schools in South Auckland (McNaughton, MacDonald, Amituanai-Toloa, Lai & Farry, 2006). Information on this project is available online at www.tlri.org.nz/project/2003/index.html#stuartmcnaughton.

The research and development programme was conducted over three years with up to 70 teachers and, in different years, between 1200 and 1900 students, over 90 percent of whom were Māori or Pasifika. Included were six Samoan bilingual classes from two schools with between 140 and 169 students across different years. A quasi-experimental design was employed to examine relationships between the programme and the outcomes over three years. The robustness of the design was enhanced by features such as a comparison with an untreated cluster of similar schools, and checks on subject attrition. Repeated measures of student achievement at the beginning and the end of each year, and a final measure at the beginning of the fourth year, form the basis of the design which, among other things, examines rates of gain against predicted patterns of growth generated from a baseline.

The initial step in the programme involved collecting baseline "profiles" of achievement, using two standardised assessments of reading comprehension It also involved collecting baseline profiles of classroom instruction, and using systematic observations in classrooms. Together these baselines provided detailed evidence about strengths and weaknesses in the students' reading comprehension, which were able to be mapped on to patterns of instruction in the classroom. For example, it showed that low decoding levels were generally not a problem; rather, it was patterns of checking and detecting threats to meaning in paragraph comprehension, and size and knowledge of vocabulary, that were posing difficulties. An unpredicted finding was that while high rates of explicit strategy instruction occurred, students were focused on the strategies as ends in themselves, and often resorted to guessing. Classroom observations showed a low incidence of teachers or students monitoring and checking strategies, and low rates of identifying and elaborating meanings of low-frequency words, unusual uses of common words, or idiomatic uses. Detailed information on the findings from the baseline is contained in Lai, McNaughton, MacDonald & Farry (2004).

The first phase included systematic feedback and analysis and problem solving at cluster, school, and classroom levels, using the profiles as evidence. This process involved two key steps. Firstly, a close examination of students' strengths and weaknesses and of current instruction to understand learning and teaching needs and secondly raising competing theories of the 'problem' and evaluating the evidence for these competing theories. This process ensured that the collaboration was a critical examination of practice and that valid inferences were drawn from the

information. The feedback procedures with examples are described fully in Lai and McNaughton (2008) and Robinson and Lai (2006). A second phase added targeted professional development, based on the evidence in the first phase, with all the Year 4–9 teachers. The professional development was designed using the profiles and known dimensions of effective teaching, and the curriculum for the sessions used a mixture of theoretical and research based ideas as well as teacher investigation and exemplification from their own classrooms. The third phase involved planned sustainability of the professional learning communities, with teacher-designed projects and a cluster-led conference.

At baseline, students were on average at stanine 3.1, approximately two years below expected levels, and this was generally the case, with some variation across year levels and across schools. To test the impact of the programme, a number of different analyses were made using longitudinal cohorts, comparisons with baseline projections, and total school population changes.

Analysis of achievement for longitudinal cohorts showed substantial acceleration had occurred and that by the end of the project, 71% of students were now in middle to upper bands of reading comprehension for their age level compared with only 40% at the start. Seventy seven percent of students would be expected in these bands nationally, indicating an educationally significant achievement given the long standing nature of the challenge to effective instruction. The average student now scored in the average band of achievement (stanine 4.21). The level of gains overall were in the order of one year's gain in addition to nationally expected progress over three years. The overall effect size for gains in stanines was 0.62. Effect sizes reported internationally are between 0.1 and 0.3 for interventions running under six years (Borman, 2005). Māori students' achievement accelerated at similar rates to those of the other ethnic groups, so that by the end of the project, the average Māori student scored within the average band (mean =4.73), with one cohort of Māori students (Year 4) scoring above the national expected average at stanine 5.29. Males and females made similar rates of progress over the three years in the intervention, but female students, on average, started with higher levels of achievement than male students. On average, students in each school made accelerated gains in achievement from the beginning to the end of the project.

Analyses using the design format showed that after two years and after three years, students had statistically significantly higher achievement than baseline comparison groups (effect sizes

ranged between 0.31 and 0.59), and were achieving statistically significantly higher than a comparison cluster of schools (effect sizes ranged between 0.33 and 0.61.) By way of comparison, effect sizes reported internationally are between 0.1 and 0.3 for interventions running under six years (Borman, 2005).

When total school populations were analysed (which included new students entering and students leaving), a similar picture to that of the previous analyses emerged. The overall level of achievement showed a variable but increasing trend over time, so that by the end of the intervention, the average stanine for 1700 students at 7 schools was 3.61. A range of gains was made between schools and within schools across the three phases. Several factors were suggested as contributing to these differences in gains, including degree of participation by schools and teachers, and aspects of curriculum planning.

The analyses of students in Samoan bilingual classrooms showed that the programme was effective in those classes too. Gains by students in the bilingual classrooms were at least as high as the gains by Samoan students in the mainstream classrooms, and in three of the year levels, they were noticeably higher. Students in bilingual classrooms were significantly lower in English reading achievement in Year 4 and Year 5, but from Year 6 onwards, their achievement levels in English were similar. Overall, cohorts made 0.8 stanine gain in two years; for four cohorts, this was a higher rate of gain than for Samoan students in mainstream classes. Gains in these classrooms could also be linked with the degree of participation by schools and teachers.

The improvements are even more significant considering the national picture of improvements in these age bands. A recent study examining trends in national data bases for students of the same age range as those reported here reveals that nationally, scores in reading comprehension have remained relatively stable for many years despite substantial changes in oral reading accuracy (Elley, 2005). A recent national review of all government funded schooling improvement initiatives further indicates that there was only initiative that has been able to improve achievement for serving culturally and linguistically diverse communities, and that is the initiative we are reporting on here (Annan, 2007).

The analyses suggest that thinking about and critically discussing the evidence at a classroom, school, and cluster level led to a significant part of the overall gains in achievement and that the

professional learning communities had the capacity to use the evidence to make changes to existing practices. This is consistent with the research linking similar processes to improving achievement (e.g., Cawelti & Protheroe, 2001), and recent reviews identifying problem-solving around evidence gathered from one's own school as an effective form of professional development (Timperley et al., 2008). What this suggests is that in general the teachers through the professional communities within schools had the capacity to change practices but needed support to identify the locus of change and test their theories about raising achievement. Given the close collaboration with researchers and policy-makers, this also confirms the importance of external support in particular research-practice-policy collaborations (e.g., Annan, 2007).

The analyses of instruction showed that specific aspects of instruction changed, including the focus on checking and detecting threats to gaining meaning in texts and boosting vocabulary acquisition, consistent with the focus of the programme and consistent with the gains that were made. But they indicated the need for caution in making assumptions about instructional and learning needs from the existing literature alone. They also indicated that effective instruction needed to be designed to fit the context-specific needs created by past histories of schooling and contemporary profiles. Interestingly, gains on decoding also increased to about the same degree as gains in other areas, despite not being a direct target of the intervention. The educational intervention also impacted on Samoan students' achievement in bilingual classrooms, demonstrating that Samoan students in bilingual classes can develop literacy in English to levels similar to those of other Samoan students who are not in bilingual classes. The evidence showed that developmental changes in English comprehension came to reach mainstream levels by around Year 6, but that this rate of change may be modifiable too. It is important to see these results in a wider developmental and educational context, involving bilingual and biliteracy development in these classes.

Observations of classroom instruction were carried out systematically in both the first and the second years. Significant changes in types of teacher and student exchanges relating to the focus of the intervention were linked to the pattern of the gains over two years in the component tests. Further case studies of teachers showed that a high gain teacher more often directed students' awareness to the requirements of activities, clarified her high expectations, pushed her students with complex tasks, introduced more complex and less familiar language including idiomatic uses, created a classroom community that enjoyed the use and study of oral and written language,

exposed students regularly to rich and varied texts, and was able to incorporate student cultural and linguistic resources, as well as clarifying areas of confusion.

We concluded that it is possible to develop more effective teaching that impacts directly on the reading comprehension achievement of Year 4-9 children. The level of gains overall were in the order of one year's gain in addition to nationally expected progress over three years. When these gains are considered in terms of the history of schooling in South Auckland, the educational significance of the gains, and the international literature of schooling improvement, they are seen to be substantial. Even when results for all the students present from the beginning to the end are considered, including those who subsequently left and those who subsequently entered the school, either from earlier levels or as new students from other schools, the levels of achievement at the schools have increased considerably. Given the quasi-experimental design with its additional strengths, these gains can be attributed with some confidence to the effects of the three-phase model adopted by the research and development programme.

Replication across clusters

The tests of effectiveness for this initial study were achieving accelerated rates of achievement and shifting distributions of achievement to match national expectations. The former sets the test at being about making more than just a normal rate of progress because that means perhaps higher levels but parallel tracks of achievement. The latter sets the test as achievement for students in the schools being no different from the distribution of the achievement for students nationally (i.e., the same proportions of low, middle and high achieving students). The first study showed that substantial acceleration was possible and significant changes in the distribution of achievement could occur, although to fully match the nationally expected distribution continued acceleration was needed.

This initial study of the process of change has been replicated twice; each time with similar results. The replication sequence which can be considered as a test of scalability has used the model developed in a first cluster of schools (Mangere) and tested it in a like cluster of schools (Otara) and in an unlike cluster of schools (West Coast). The like set of schools were from an adjacent neighborhood to the first and whose students were from the same communities of Māori (indigenous) and Pasifika (Minority) communities with the lowest income levels and starting

achievement about two years below national expectations. The unlike set of schools were from a small town and rural area of New Zealand and involved mainly NZ European and Māori students in communities with higher income levels and starting achievement levels around national expectations. All in all, the replication sequence involved 48 schools, representing about 7000 students yearly. We statistically modeled the data in each cluster to predict the amount of gain for each year of the intervention.

The implementation of the model in all three settings raised achievement every phase of the intervention and the average rate of gain in each phase of the intervention was 0.32 for the West Coast (range of gain between 0.27 and 0.38 stanine), 0.26 for Otara (range of gain between 0.18 and 0.33) and 0.30 for Mangere (range of gain between 0.24 and 0.36) (See Figure 1 for a visual representation of the mean rate of gain and the confidence intervals). This means that in every phase across all three clusters, students in the intervention gained about three months in addition to nationally expected progress over the one year period.

Figure 1. The estimated effects of each intervention plotted along with respective 95% confidence intervals at the three clusters of schools
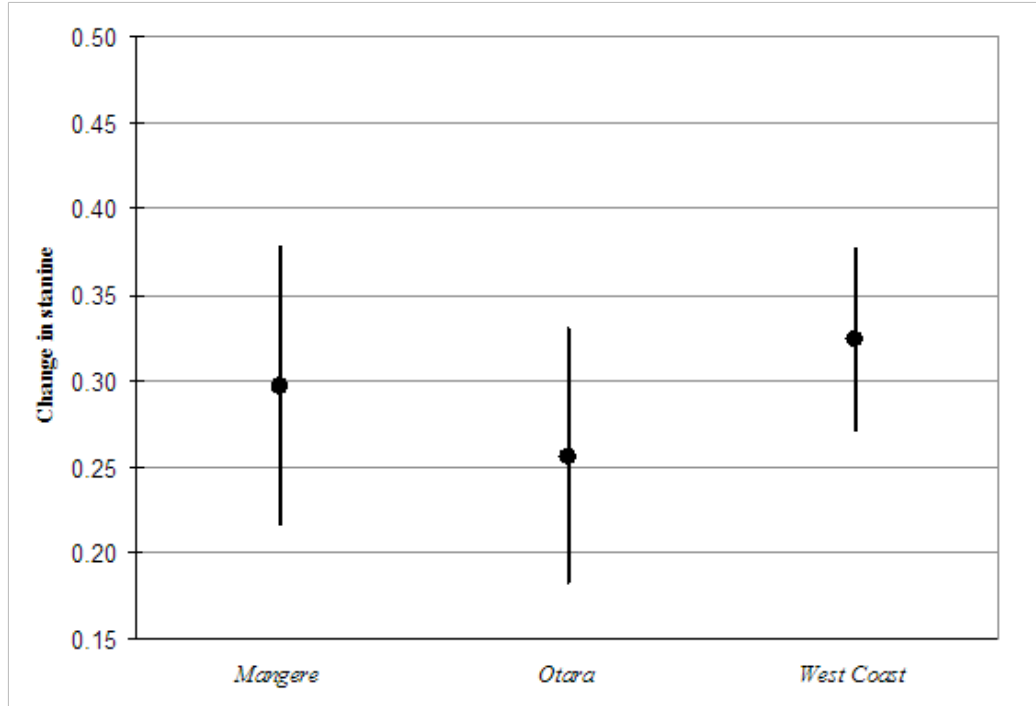

The results also indicated that the intervention produced the most gain for the West Coast cluster. However, the intervention at the West Coast was most affected by the summer holidays, in that

there was an estimated average of 0.25 stanine drop over every summer holiday during the three year program, with the range of the drop in stanine in each summer ranging from 0.19 to 0.31 stanine. Summer effects were not significant at both Mangere and Otara clusters. This means that whilst students in the West Coast made more gain that the other clusters during the academic year, students were less likely to retain their learning over the summer holidays and their results went down from the end of the academic year to the beginning of the following one. Conversely, whilst the Mangere and Otara clusters made less gain during the academic year, their scores did not go down significantly from the end of the academic year to the beginning of the following year. Figure 2 shows these patterns.

Thus the results suggest that the findings from the first study are able to be replicated and scaled up across different settings, different schools, and different cohorts of students with a variety of starting achievement levels.

Figure 2. Models of the estimated progressions of reading achievements over three years in three clusters of schools.
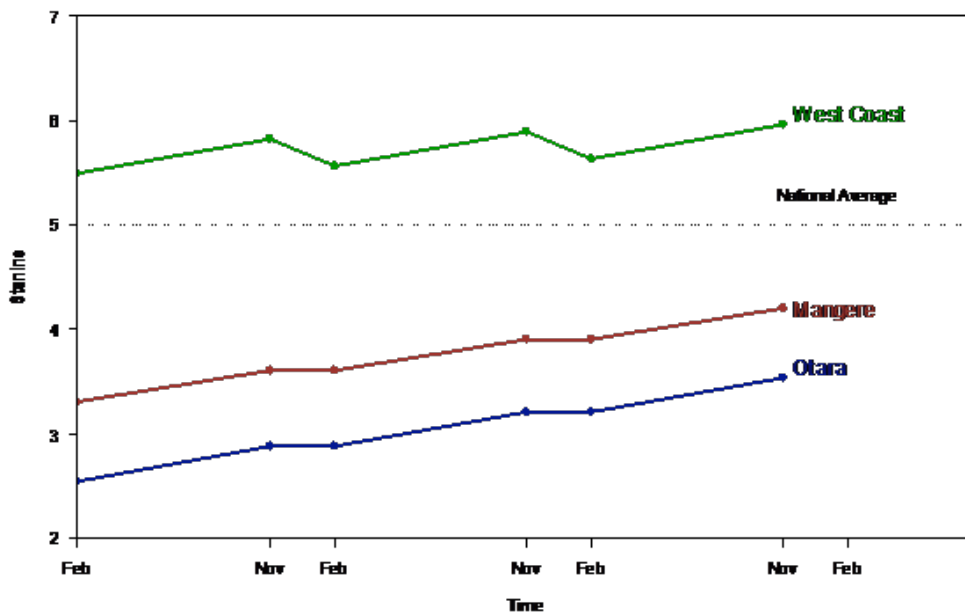
References

Annan, B. (2007). *A theory for schooling improvement: Consistency and connectivity to improve instructional practice.* Unpublished doctoral thesis, University of Auckland, Auckland, New Zealand.

Alton-Lee, A. (2004, April). *Improving educational policy and practice through an iterative best evidence synthesis programme.* Paper presented at the OECD-US seminar, Evidence-ased Policy Research, Washington, D.C.

Bandura, A. (1995). Exercise of personal and collective self efficacy in changing societies. In A. Bandura (Ed.), *Self efficacy in changing societies* (pp. 1-45). Cambridge: Cambridge University Press.

Buly, M. R., & Valencia, B. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, *24*(3), 219-239.

Block, C. C., & Pressley, M., (Eds.). (2002). *Comprehension Instruction: Research-based best practices.* New York: Guildford Press.

Borman, G. D. (2005). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. In L. Parker (Ed.), *Review of Research in Education, 29* (pp. 1-28). Washington, DC: American Educational Research Association.

Cawelti, G., & Protheroe, N, (2001). *High student achievement: How six school districts changed into high-performance systems.* Arlington, VA: Educational Research Service.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3-12.

Cohen, D.K., & Ball, D.L. (2007). Educational innovation and the problem of scale. In B. Schneider & S. McDonald (Eds.), *Scale up in education: Ideas in principle (*Vol. 1) (pp. 19-36). Lanham, MD: Rowman and Littlefield.

Crooks, T., & Flockton, L (2005). *Reading and speaking assessment results 2004. National education monitoring report 34.* Dunedin: Educational Assessment Research Unit.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk (JESPAR), 5*(1&2), 183-204.

Darling-Hammond, L., & Bransford, J. (Eds.) (2005). *Preparing teachers for a changing world.* San Francisco: John Wiley.

Davis, B. & Sumara, D. (2003). Why aren't they getting this? Working through the regressive myths of constructivist pedagogy. *Teaching Education, 14*(2) 123-140.

Dyson, A. H. (1999). Transforming transfer: Unruly students, contrary texts and the persistence of the pedagogical order. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education, 24* (pp. 141-171). Washington, DC: American Educational Research Association.

Elley, W. (2005). On the remarkable stability of student achievement standards over time. *New Zealand Journal of Educational Studies*, *40*(1) 3-23.

Guthrie, J.T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.). *Handbook of reading research: Volume III* (pp. 403-422). New York: Erlbaum.

Hubbard, L., Mehan, B., & Stein, M. K. (2006). Reform as learning : School reform, organizational culture, and community politics in San Diego. New York : Routledge.

Lai, M. K., McNaughton, S., MacDonald, S., & Farry, S. (2004). Profiling reading comprehension in Mangere schools: A research and development collaboration. *New Zealand Journal of Educational Studies, 27*(3), 184 – 197.

Lai, M.K., & McNaughton, S. (2008). Raising student achievement in poor, urban communities through evidence-based conversations. In L. Earl & H. Timperley (Eds.), *Professional Learning Conversations: Challenges in Using Evidence* (pp.13-27). Netherlands: Kluwer/Springer Academic Publishers.

Lave, J. (1991). Situated learning in communities of practice. In L. B. Resnick, J. M. Levine & S.

    D. Teasley (Eds.), *Perspectives on socially shared cognition.* (pp.63-82) Washington,

    DC: American Psychological Association.

Lefstein, A. (2007). *Changing teacher practice through the National Literacy Strategy: A*

    micro-interactional perspective. Unpublished manuscript, Department of Educational

    Studies, Oxford University.

Louis, K. S., Marks, H., & Kruse, S. (1996). Teachers' professional community in restructuring

    schools. *American Educational Research Journal, 33*(4), 757–798.

McCall, R. G., & Green, B. L. (2004). Beyond the methodological gold standards of behavioural

    research: Considerations for practice and policy. In L. Sherod (Ed.), *Social Policy Report.*

    Giving Child and Youth Development Knowledge Away. Vol. XVIII, No. II. Ann Arbor,

    MI: Society for Research in Child Development.

McNaughton, S. (2002). *Meeting of Minds*. Wellington: Learning Media.

McNaughton, S., MacDonald, S., Amituanai-Toloa, M., Lai, M., & Farry, S. (2006). *Enhanced*

    teaching and learning of comprehension in Years 4-9 in seven Mangere schools: Final

    TLRI report. Wellington: NZCER. Retrieved June 30, 2006, from http://www.tlri.org.nz/

    pdfs/9206_finalreport.pdf

New London Group (1996). A pedagogy of multiliteracies: Designing social features. *Harvard*

    Educational Review, 66, 60-92.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research*

    Quarterly, 40(2), 184-202.

Phillips, G., McNaughton, S., & MacDonald, S (2004). Managing the Mismatch: Enhancing

    early literacy progress for children with diverse language and cultural identities in

    mainstream urban schools in New Zealand. *Journal of Educational Psychology, 96*(2),

    309-323.

Ramsay, P. D. K., Sneddon, D. G., Grenfell, J., & Ford, I. (1981). *Tomorrow may be too late.*

Final report of the schools with special needs project. Hamilton, New Zealand: Education Department, University of Waikato.

Robinson, V., & Lai, M. K. (2006). *Practitioner research for educators: A guide to improving classrooms and schools.* Thousand Oaks, CA: Corwin Press.

Robinson, V. M. J., Lloyd, C. & Hohepa, M. (2007, April). *The effect of leadership on students.* Paper presented at the American Education Research Association Conference, Chicago, IL.

Rowan, B., Camburn, E. M., Correnti, R., & Miller, R. (2007). *How comprehensive school reform works: Insights from 'A Study of Instructional Improvement'.* Unpublished report by the Consortium for Policy Research in Education as part of 'A Study of Instructional Improvement (SII)'. Provided by the first author.

Seashore-Louis, K. (2006). Changing the culture of schools: Professional community, organizational learning, and trust., *Journal of School Leadership, 16,* 477-489.

Slavin, R,. & Madden, N. (2001). *One million children: Success for all*. Thousand Oaks, CA: Corwin.

Taylor, B.M., Pearson, P.D., Peterson, D., & Rodriguez, M.C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40*(1), 40-69.

Timperley, H., Robinson, V. M. J., & Bullard, T. (1999). *Strengthening education in Mangere and Otara: final evaluation report*. Wellington: Auckland Uniservices Ltd for the Ministry of Education.

Timperley, H., Phillips, G., & Wiseman, J. (2003). *The Sustainability of Professional Development in Literacy Parts One and Two*. Wellington: Auckland Uniservices Ltd for the Ministry of Education.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2008). *Best evidence synthesis iteration (BES) on professional learning and development.* Wellington: Ministry of Education.

Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*.

Cambridge: Cambridge University Press.