

(Using) evidence for educational improvement

Adrienne Alton-Lee*

Iterative Best Evidence Synthesis Programme, Ministry of Education, Wellington, New Zealand

(Received 15 February 2011; final version received 7 July 2011)

The New Zealand Ministry of Education has published a best evidence synthesis iteration (BES) that identifies the characteristics of teacher professional development that make a positive difference for valued student outcomes. A companion best evidence synthesis iteration (BES) that identifies the leadership influences on valued student outcomes reveals that when school leaders promote and/or participate in effective teacher professional learning and development, this has more impact on student achievement than any other leadership activity. This article provides an overview of the findings about effective professional development and highlights the potential of such evidence to inform educational improvement. An example of a tool to support collaborative professional inquiry and knowledge building in schools is provided. The article highlights policy challenges for the systemic use of effective professional development and illustrates what is possible in two examples of high impact research and development (R & D) that have been effective across varied contexts. These examples illustrate the potential for educational improvement when professional learning is underpinned by cumulative high impact research and development in education.

Keywords: leadership; professional development; student outcomes; social outcomes; academic achievement; evidence; policy; equity; improvement

There has been a silent revolution in education policy around the world in recent times. Historically, most education systems were charged with sorting students by achievement level, into those who succeeded and those who failed. In some countries, such sorting has been a feature of schooling even at the points of access to primary and/or secondary education.

As knowledge itself is increasingly seen as an economic resource in a global community, societies are looking for school systems that help all students to learn, succeed and develop the capabilities needed to be lifelong learners in rapidly changing contexts. Increasingly diverse societies are also looking to school systems to contribute to social cohesion. This silent revolution requires school systems to perform at much higher levels, particularly for those who have traditionally been underserved by education (Bishop & Glynn, 1999), and/or educationally disadvantaged by socio-economic inequality (Ainscow et al., 2010; Levin, 2009).

That the task of schools has undergone a fundamental change; from sorting, to supporting all learners to succeed, needs to be acknowledged in policy

*Email: Adrienne.Alton-Lee@minedu.govt.nz

environments. Policy ideas for responding to this change are likely to be superficial if they do not attend to the implications for capability building. If the magnitude of the change required goes unrecognised, there is a risk of policy discourses being almost magical in their assumptions about how change might be brought about. For example, it may be assumed that all that needs to happen is for teachers – working with large groups of students – to ‘personalise’ learning for each child. This article calls for much greater attention to be paid to the role of research and development in informing professional learning and systemic capability building and for less faith to be placed in the ‘teachers must try harder’ or magical discourses that can pervade policy thinking.

For some academics, attempting educational improvement and equity through schooling is futile because, they argue, schooling can make little difference to relative life chances of students from different socio-economic status families: ‘The much lamented “tail” of achievement is the tail of poverty. It is perverse to suggest that this can be removed by changing teachers’ attitudes and providing different approaches to learning’ (Snook, 2008, pp. 5–6).

For New Zealand, the issue of the influence of socio-economic background on educational outcomes is a critical policy issue. The findings from the 2009 PISA reading performance show that ‘some countries have been much more successful than others in mitigating the impact of socio-economic status on student achievement’ (OECD, 2010, p. 48). New Zealand was least successful out of 65 participating countries in mitigating the influence of the socio-economic status of students’ families on their reading achievement.

The Iterative BES Programme commissioned an early synthesis of evidence about the impact of families and communities on educational outcomes that has informed advice about the educational impact of factors such as poverty, family resources such as books and computers, child hearing loss, television viewing and the quality of interactions at home (Biddulph, Biddulph, & Biddulph, 2003). Clearly such factors make a difference and such evidence has significant implications for broader policy decisions. Levin (2009) points out, however:

We have learned that while changing classroom practice is difficult, and by itself will not be enough, significant improvement in outcomes for poor children cannot occur unless there is change in teaching and learning practices. (Levin, 2009, p. 191)

Harker (2007) found that the interaction of English-medium schooling with Māori and Pasifika student ethnicity in New Zealand had a negative impact that is not explained by socio-economic status of families or school-mix; a finding that is consistent with evidence about the centrality of culture in teaching and learning (Alton-Lee, 2003). Levin warns of the fiscal consequences of convincing policy-makers that schools cannot make a difference to inequalities of outcome:

This debate is important because it has much to say to governments about where they should invest scarce resources. If schools are a prime agency for addressing inequalities then that is where additional energy and money should go. If, however, schools are less important than other areas of social policy such as housing or employment or early childhood, then it follows that the resources should be allocated to these other areas, possibly even being reallocated away from schools. (Levin, 2009, p. 191)

This article focuses on the findings of a series of best evidence syntheses that reveal how it is possible to make a transformational difference to valued student outcomes through systemic improvement in schooling. I first make the case that there is an emerging and compelling evidence base for the role that effective teacher professional learning and development has to play in resourcing and enabling improvement in education. The evidence provided is derived from syntheses of evidence about influences on change and valued outcomes for diverse learners in education – academic, social, well-being, self-regulatory and identity outcomes. The history of much educational research is that attention to the link between processes *and outcomes for students* (rather than the perspectives of providers or researchers) is rare. I consider briefly the evidence of unintended harm in education. I go on to argue that research and development is not only a fruitful approach but also a moral imperative, in that it is a means to counter unintended harm and a means by which we can strengthen outcomes that our communities seek for their children. Finding what works in one setting does not in itself demonstrate what is needed to spread such reform more widely. The policy challenge of scaling-up effective professional development across a whole school system is a recurring theme in the article.

Our best evidence synthesis methodology has enabled us to analyse the comparative size of effect of different approaches and interventions. In this article I describe two outstanding collaborative research and development (R & D) programmes: *Project Fast* and *Complex Instruction*. Both are high-impact professional development interventions underpinned and refined through cycles of R & D that have advanced multiple valued outcomes for students across very different settings.

The article includes consideration of student cooperation during group work – a neglected policy focus for improving academic and social outcomes; a research area in which Maurice Galton's early contribution to the field has been seminal. I echo Maurice Galton's ongoing calls for action around systemic use of such evidence about making a bigger difference.

This article counterpoints OECD (2003, 2007) findings of relatively low expenditure on research and development in education with the value of collaborative, outcomes-focused R & D as a resource for effective policy and practice. Given the demonstrated potential of outcomes-focused R & D and its capacity to counter the risks of unintended harm in education, an argument is made for strategic use of R & D to inform professional learning in ways that can lift the capability and performance of schooling.

Evidence from educational leadership research

In considering the evidence for the importance of teacher professional development, I draw initially on the very small subset of research in the school leadership literature that focuses on student outcomes. The role of teacher professional learning has been highlighted in an analysis of the relationship between school leadership activities and student outcome gains prepared for the New Zealand Ministry of Education's *Educational leadership and student outcomes; Identifying what works and why best evidence synthesis BES* (Robinson, Hohepa, & Lloyd, 2009; see Figure 1).

The number of effects and studies considered and the standard errors for these effects are elaborated further in Table 1 over. The source studies are cited in

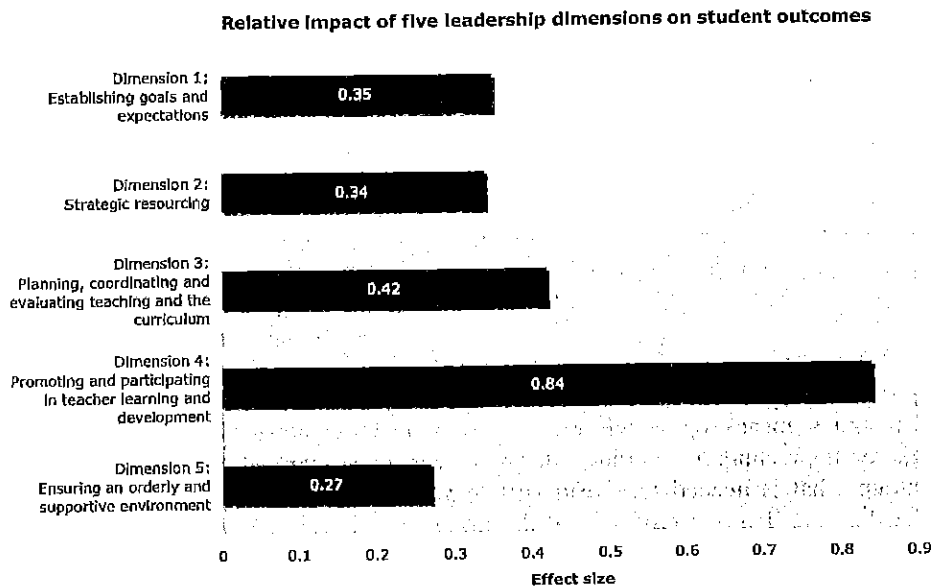


Figure 1. Relative impact of five leadership dimensions on student outcomes.

Robinson, Hohepa and Lloyd (2009) and the analysis is also available in Robinson, Lloyd and Rowe (2008).

By far the greatest effect size is associated with the role of school leaders in promoting and participating in teacher professional learning and development ($d=0.84$). This leadership practice has twice the impact on student outcomes of any other leadership focus. Further analysis suggests that, by doing this, school leaders not only deepen their own pedagogical knowledge and understanding but they also develop the understandings necessary to create and sustain the conditions for improved practice in their schools.

The constrained nature of the available empirical evidence did not allow a comparable quantitative analysis of three other leadership dimensions found to be significantly linked to valued student outcomes. These findings were leadership capacity to: create educationally powerful connections (particularly with the experiences and families of learners of indigenous and/or diverse cultural identities and families of low socio-economic status), engage in constructive problem talk, and select, develop and use smart tools. Four further findings about leader knowledge, skills and dispositions linked to improved student outcomes were leadership capacity to: ensure administrative decisions are informed by knowledge about effective pedagogy, analyse and solve complex problems, build relational trust and engage in open-to-learning conversations. Each of these findings is explained fully in the source best evidence synthesis which is available online. Case 5 of that BES reveals how a school leadership team were able to establish and sustain a four workshop programme with families that enabled parents to support their children's reading and gain access to books through a relationship with a local library. This intervention resulted in an achievement difference of more than a year's reading achievement gain compared with a control group.

Table 1. Leadership Dimensions Derived from Studies of Effects of Leadership on Student Outcomes.

Leadership Dimension	Meaning of Dimension	Effect Sizes/No. of studies
1. Establishing Goals and Expectations	Includes the setting, communicating and monitoring of learning goals, standards and expectations, and the involvement of staff and others in the process so that there is clarity and consensus about goals.	Average ES = 0.35 (SE=0.08) 49 effect sizes from 7 studies
2. Strategic Resourcing	Involves aligning resource selection and allocation to priority teaching goals. Includes provision of appropriate expertise through staff recruitment.	Average ES = 0.34 (SE=0.09) 11 effect sizes from 7 studies
3. Planning, Coordinating and Evaluating Teaching and the Curriculum	Direct involvement in the support and evaluation of teaching through regular classroom visits and provision of formative and summative feedback to teachers. Direct oversight of curriculum through school-wide coordination across classes and year levels and alignment to school goals.	Average ES = 0.42 (SE=0.07) 79 effect sizes from 7 studies
4. Promoting and Participating in Teacher Learning and Development	Leadership that not only promotes but directly participates with teachers in formal or informal professional learning.	Average ES = 0.84 (SE=0.14) 17 effect sizes from 6 studies
5. Ensuring an Orderly and Supportive Environment	Protecting time for teaching and learning by reducing external pressures and interruptions and establishing an orderly and supportive environment both inside and outside classrooms.	Average ES = 0.27 (SE=0.09) 42 effect sizes from 8 studies

Implications for leadership

The *School leadership and Teacher professional learning and development BES* findings call for school leadership to play a central role in embedding a professional inquiry model into teaching practice. For example, the findings highlight how important it is for school leaders to actively develop shared commitment to goals that involve improving student outcomes, to promote and lead professional development and to manage constructive problem talk. Effectiveness is linked to the role of leadership in creating and sustaining the conditions for ongoing, outcomes-focused professional inquiry and learning in schools. Such conditions include enabling teachers to process new learning with others and providing multiple opportunities for teachers to learn and apply their new understandings in practice.

A recurrent finding in the emerging literature about the use of evidence is that use is more likely when there is stakeholder ownership. In a self-managing school system, the ownership of the profession matters. The Chair of the New Zealand Secondary Principals' Council wrote:

The Leadership BES already has significant traction in New Zealand secondary schools and is well regarded by the profession as being both aspirational and practical in content. We are proud to have been involved with it from the beginning and

commend it to you as a well researched, clear, and detailed way forward for leaders at any level of the schooling system. We hope it gives principals in particular a focus for their work as well as being a useful tool to help us all find ways to improve student outcomes. (Davidson, 2011)

International formative quality assessors of this BES, Professors Ben Levin and Michael Fullan warn that irrespective of how impressive and valuable the BES, its potential as a tool for systemic improvement depends on the policy response to its 'huge implementation challenge'. They explain that 'This BES report on leadership will be for nought unless there is a concerted plan' around a 'concrete problem' such as raising achievement and reducing disparity that 'incorporates the key leadership capacities into the implementation of that plan' (Levin & Fullan, 2009, p. 15).

Evidence about the effectiveness of teacher professional development

The second body of evidence informing this paper is derived from a synthesis (of findings from 72 individual or groups of studies with 227 effects) that analysed the links between professional development and its impact on valued student outcomes (Timperley, Wilson, Barrar, & Fung, 2007). Table 2 provides an overview of the findings of one of the quantitative analyses that informed the synthesis. There was much variability in effectiveness by curriculum area, with negative effects as well as positive effects occurring. The highest effect sizes tended to occur for low achievers and students with special needs whose teachers participated in very effective professional development. Improvements for academic outcomes were, on average, larger than those achieved for social or other outcomes.

In giving guidance on interpreting effect sizes Cohen (1988) suggested an effect of $d=0.20$ can be considered small, an effect of $d=0.50$ medium and an effect of $d=0.80$ large. When interpreting the table, it is important to realise that the size of the effect is not the only criterion by which to compare interventions; length is another; some of the shorter interventions had significant impact relative to their duration. In general, professional development that continued in some form – often after intensive teacher engagement for one to two years – was found to have the greatest impact on student outcomes.

Another way of judging the effectiveness of professional development is to compare the effect sizes for student gains with those achieved under conditions of business-as-usual. By analysing standardised asTTle (Assessment Tools for Teaching and Learning) assessments, Hattie (2009) has found that, in general, the effect size for a New Zealand teacher's contribution to student learning over a year is around $d=0.35$:

In our own New Zealand studies, we have estimated the yearly effect in reading, mathematics, and writing from Years 4 to 13 ($N=83,751$) is $d=0.35$, although this is not linear: in some years and for some subjects there is more or less growth. The inference for the argument in this book is that teachers typically can attain between $d=0.20$ to $d=0.40$ growth per year, and that this is to be considered *average*. They should be seeking greater than $d=0.40$ for their achievement gains to be considered *above average*, and greater than $d=0.60$ to be considered *excellent*. (Hattie, 2009, p. 17)

Table 2. The Range and Mean Effect Sizes for 72 Studies of Impact of Professional Development on Student Outcomes (Sources: Timperley, Wilson, Barrar & Fung, 2007; Timperley & Alton-Lee, 2008).

All effects								
	No. of studies	Mean effect size	Standard Error of the mean	95% Confidence interval	Median	Standard deviation	Minimum effect size	Maximum effect size
Total	227	0.60	0.06	0.24	0.34	0.83	-1.01	5.31
<i>OUTCOME</i>	F=3.30; p=.001; $\eta^2=.13$							
Mathematics	62	0.50	0.12	0.48	0.31	0.94	-1.01	5.10
Reading	44	0.34	0.04	0.16	0.26	0.26	-0.01	1.11
Literacy / Language Skills	27	1.18	0.24	0.96	0.55	1.27	0.09	5.31
Attitudes toward Subject	21	0.34	0.21	0.84	0.11	0.95	-0.73	4.27
Science	18	0.94	0.19	0.76	0.68	0.80	0.16	2.85
Writing	16	0.88	0.11	0.44	1.06	0.45	0.06	1.34
Self-Efficacy	11	0.17	0.06	0.24	0.11	0.21	-0.07	0.68
Other Academic Skills	10	0.76	0.18	0.72	0.55	0.57	0.22	2.09
Social Outcomes	7	0.36	0.11	0.44	0.34	0.29	-0.11	0.86
Cognitive Processing	6	0.85	0.18	0.72	0.87	0.44	0.17	1.46
Other Personal Outcomes	5	0.46	0.10	0.40	0.53	0.23	0.08	0.64
<i>Class of Outcome</i>	F=3.25; p=.041; $\eta^2=.03$							
Academic	183	0.66	0.06	0.24	0.39	0.85	-1.01	5.31
Personal	37	0.30	0.12	0.48	0.12	0.73	-0.73	4.27
Social	7	0.36	0.11	0.44	0.34	0.29	-0.11	0.86
<i>Grade Level Groupings</i>	Ns							
Elementary	172	0.61	0.07	0.28	0.34	0.90	-1.01	5.31
Junior High	23	0.36	0.06	0.24	0.27	0.30	0.05	1.27

(Continued)

Table 2. (Continued)

All effects									
	No. of studies	Mean effect size	Standard Error of the mean	95% Confidence interval	Median	Standard deviation	Minimum effect size	Maximum effect size	
Secondary ALL	20	0.60	0.14	0.56	0.45	0.61	0.06	2.85	
	9	0.97	0.32	1.28	0.64	0.95	0.08	2.68	
<i>Country</i>	Ns								
United States	143	0.48	0.07	0.28	0.27	0.80	-1.01	5.10	
New Zealand	68	0.87	0.11	0.44	0.53	0.90	-0.14	5.31	
Canada	4	0.79	0.44	1.76	0.43	0.88	0.23	2.09	
The Netherlands	4	0.48	0.22	0.88	0.36	0.44	0.09	1.12	
United Kingdom	4	0.53	0.13	0.52	0.49	0.27	0.29	0.85	
Israel	2	0.26	0.01	0.04	0.26	0.01	0.25	0.26	
other country	1	0.31			0.31		0.31	0.31	
<i>Number of Participants</i>	Ns								
<100	20	0.84	0.13	0.52	0.64	0.57	0.21	2.68	
100-999	83	0.69	0.11	0.44	0.42	0.96	-0.73	5.10	
>1000	56	0.69	0.13	0.52	0.32	1.00	-0.03	5.31	
<i>Type of Control</i>									
control	138	0.50	0.07	0.28	0.31	0.81	-1.01	5.10	
baseline	89	0.75	0.09	0.36	0.45	0.85	0.04	5.31	
<i>Type of Instrumentation</i>									
Objectively Scored	119	0.40	0.05	0.20	0.28	0.51	-0.14	4.27	
Researcher Verified	80	0.62	0.10	0.40	0.38	0.92	-1.01	5.10	
Judgment	28	1.39	0.22	0.88	1.27	1.15	0.16	5.31	

As can be seen in Table 2, many effects were significantly larger than those for business-as-usual in controlled comparisons. In some cases, the gain in one year was equivalent to three or more years' 'normal' gain, with greater gains made by the lowest achieving students. While time and resources were a necessary condition for effectiveness, there was little evidence that just providing teachers with time and resources is effective in promoting professional learning in ways that have positive outcomes for students. The nature and quality of professional learning and development matter.

First do no harm

The Iterative Best Evidence Synthesis Programme was established to bring together evidence about what works, what makes a bigger difference in education and what does not work for advancing student learning and well-being. We argue that like the Hippocratic oath in health a 'first do no harm' principle needs to be adopted in educational policy and practice.

The analysis in Table 2 (*minimum effect sizes*) reveals examples of professional development that shifted the practices of teachers in ways that had deleterious effects on student achievement and/or other outcomes. For example, there was evidence in the source studies that professional learning communities designed to address inequities could actually be counterproductive. For example, Lipman (1997) described how teachers who were given two hours of non-contact time per day to find answers to African-American underachievement interacted in ways that reinforced existing deficit thinking and structural inequalities. Those teachers who held alternative theories and could have served as a resource for the group's deliberations were marginalised. Student achievement actually declined as a result of the nature of this costly teacher professional development.

In their search for evidence, the BES writers found meagre outcomes-linked evidence available for the effects of professional development on a curriculum area that is crucial to the development of social, participatory and citizenship skills: social studies. This gap in the professional development literature is of concern not only because such outcomes are important for social cohesion, but also because there is compelling evidence from studies linking teacher goals, learning processes and student outcomes that it is possible for teachers – well-intentioned, caring and experienced – to unknowingly have impacts on students that are the direct reverse of what they intend (Aitken & Sinnema, 2008; Alton-Lee, Nuthall, & Patrick, 1995; Bossert, 1979; Doyle, 1983; Nuthall, 1999). Some efforts to combat racism, for example, have been found to exacerbate the problem (Cole, 1998; Donn & Schick, 1995; Osler & Starkey, 1999; Sexias, 2001; Shaver, 1999). Given recurrent findings of inadvertent harm done in education, there is a moral imperative for teacher educators to build on evidence about effectiveness in the design of professional development and to use ongoing evaluation of impact on student academic *and* social outcomes including identity to improve practice. For policy-makers, this means requiring systematic use of evidence about effects on student outcomes, and iterative processes to optimise effectiveness, when investing in professional development.

Characteristics of highly effective professional development

Qualitative analyses of the synthesis findings revealed the greatest gains evident were those that deepened teachers' foundation of curricula-specific pedagogical content and assessment knowledge. Such professional development provided teachers with new theoretical understandings that helped them make informed decisions about their practice.

A brief summary of the overall findings from the best evidence synthesis is available as one of the International Academy of Education's Education Practice Series on the UNESCO website (Timperley, 2008). Figure 2 highlights 10 principles of effective professional development derived from the findings.

In considering the findings about what did work reported in Figure 2 it is helpful to reflect further on the BES findings of what did not work, in order to understand policy significance. For example, when professional development bypasses rather than engages teachers' own theories about their practice, little changes for students. When teachers get limited professional development and lack opportunity to integrate their new learning into practice, little changes for students. When there is an absence of proactive involvement by leaders in the professional learning and when there are constrained or no opportunities for teachers to process new learning with colleagues, less changes for students.

These findings are important for considerations of value for investment of funding and professional time. For policy-makers seeking a rapid scale-up or a cheaper professional development option, investment and professional time can be wasted if implementation is compromised in ways that do not enable the deep change that benefits students.

For policy-makers concerned with value for investment, and educators seeking to be more effective, high impact teaching approaches matter. The BES methodology identifies examples of the highest impact professional development across source studies; including those that have been tried and tested in multiple contexts such as *Project Fast* and *Complex Instruction* (see below). This section focuses on these two professional development programmes for which there is evidence of very high impacts on student learning achieved within relatively short timeframes. Both programmes developed external expertise to support professional development that is consistent with all 10 findings of the *Teacher professional learning and development BES*. Both have evolved through decades of R & D. Both have attempted to use a cyclical model of R & D in which researchers, teacher educators, curriculum developers, practitioners, other specialists and sometimes community members, have collaborated across sites.

Project FAST (Foundational Approaches to Science Teaching)

The writers of the *Teacher professional learning and development BES* found a 2.85 effect size gain in students' achievement a year after their teachers had received professional development in *Project FAST* (Foundational Approaches to Science Teaching). For decades, evaluations of this professional development project have shown that the outcomes – scientific knowledge, thinking skills, higher-order thinking and creativity – of students of participating teachers are significantly greater than those of students from comparison groups (Young 1999). The project has repeatedly been accorded exemplary status in US reviews.

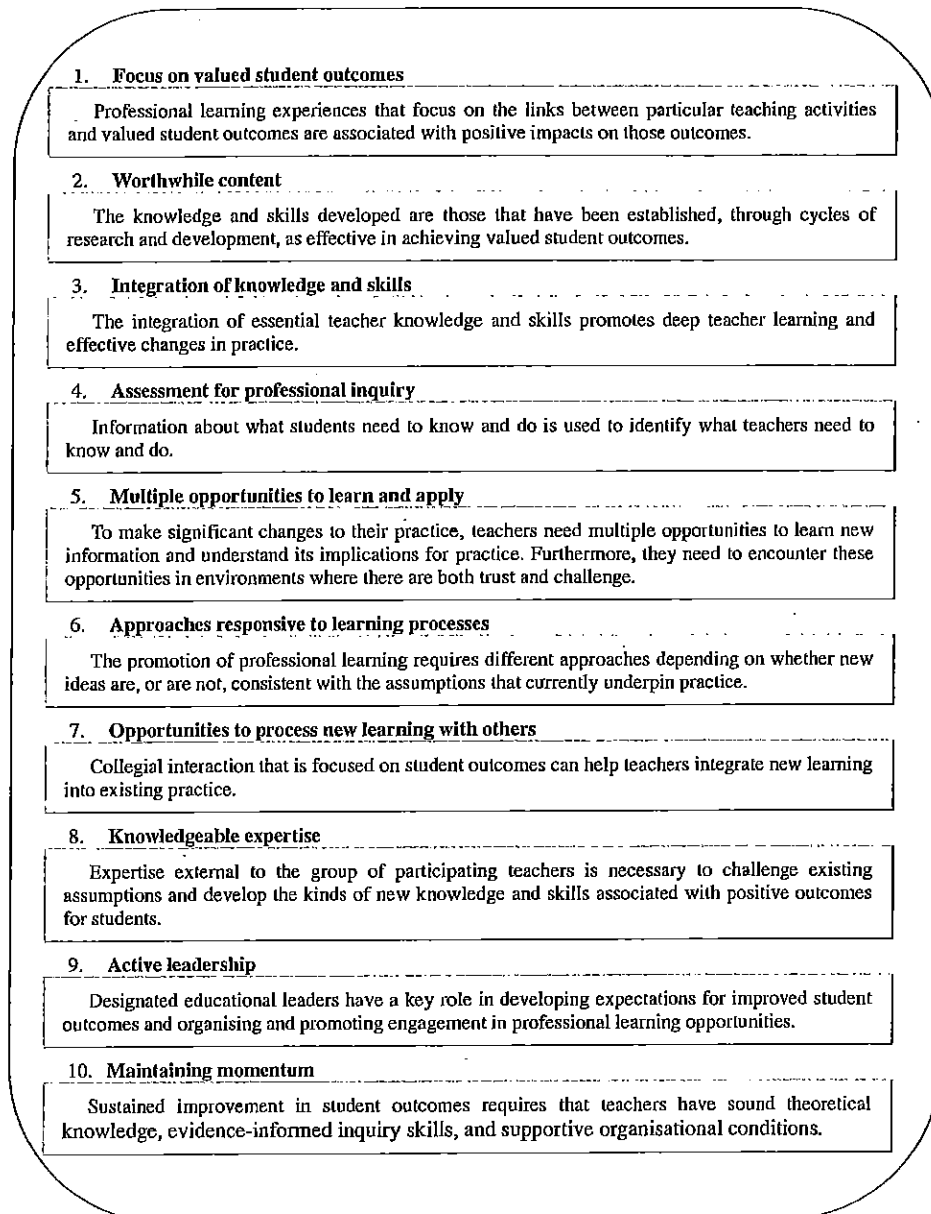


Figure 2. Summary of the findings of the Teacher Professional Learning and Development BES. Timperley, H. (2008). *Teacher Professional Learning and Development*. International Academy of Education. *Education Practice Series-18*.

Project Fast is an inquiry-based, interdisciplinary approach to science education, embracing physical science, ecology and relational study (for example, resource management, technology, air pollution, water resource management, world food production and humans in the environment). The approach is specifically designed to address the developmental needs of middle school students (12- to 15-year-olds)

and to produce equitable learning opportunities. Teachers are required to develop their students' capacity to learn cooperatively and collaboratively.

The programme and associated curriculum materials have been informed by R & D carried out by the Curriculum, Research and Development Group at the University of Hawaii since 1966. Yamamoto (2007) provided an account of the collaborative processes of research, development, trial and evaluation used by the Project Fast developers as they have worked alongside teachers in a laboratory school setting over the four decades. The magnitude of the impact of the Project Fast professional development reflects the value of repeated cycles of R & D that produced 'smart tools' and of processes that sustained an ongoing inquiry approach.

By 2001, Project Fast had been used in 11 countries, in a range of languages including Braille, and across 36 states in the US. By 2007 the National Training Foundation in Russia had collaborated with the Curriculum, Research and Development Group at the University of Hawaii to move beyond a translation of the original programme to develop a local programme and strategy for professional learning based on the FAST model. In 2009 a five-year Pacific Education and Research for Leadership in Science (PEARLS) project was established to provide professional development in FAST for teachers in the Pacific region. Recent work has been carried out to create, evaluate and refine face-to-face delivery options, which, combined with e-learning support, continue to enable substantial gains for students (Brandon et al., 2007).

Complex instruction: working for equity in heterogeneous classrooms

The writers of the *Teacher professional learning and development BES* found a 1.06 effect for students' achievement four months after their teachers had received professional development in Complex Instruction (CI).

CI is a R & D programme that was initiated by Elizabeth Cohen with Rachel Lotan at Stanford University in 1979. Cohen drew on sociological and organisational theory to develop a strategy for equitable instruction that would foster higher-order thinking in high and low achievers. The approach attended to the inherent challenge of the classroom as a social group context. Teachers are supported to use the theory to foster small-group approaches. Students are trained to value the contributions of each group member, to be accountable, and to manage cognitive dissonance constructively. Students learn to use each other as learning resources, thereby multiplying the supports available to each. Teachers are trained to use specific strategies that enhance the role of low-status students, to the benefit of both high and low achievers.

CI complements an engineering approach to group work with well-designed curricular materials that focus on big ideas and a problem-solving approach. CI strengthens teacher pedagogical content knowledge through the use of multiple-ability curricular resources consisting of carefully designed and evaluated group tasks. The collaborative R & D process draws on a wide range of subject matter, research expertise and community funds of knowledge to develop effective tasks. In the United States, bilingual instructions were normalised to support the engagement of students whose first languages are English or Spanish while also creating a learning environment that was supportive of students who speak other languages. Neves (2007) found that, in CI, increased proficiency in English was associated with more frequent use by students of their first language, Spanish.

Early studies showed positive correlations between increased rates of student talking and working together and student gains on standardised tests of mathematics, as well as on social studies tests (Cohen, Lotan, & Hothuis, 1997). Cohen et al. (1997) published evidence gathered over a 12-year period demonstrating that the students of CI-trained teachers achieved more highly on standardised tests and unit tests in social studies, science and mathematics at both elementary and middle school levels. Filby (1997) attributed the sustained use of CI by teachers in Arizona not only to improved achievement outcomes but to other outcomes such as improved motivation and reduced problems with discipline and truancy. A study of the sustainability of CI found that classroom teachers were still using the approach three years after external expertise was withdrawn because, over and above the achievement gains, both they and their students enjoyed the approach (Dahl, 1997).

The cycles of R & D used in CI have informed the quality and effectiveness of the approach at both the classroom and professional development levels. Like Project Fast, CI has generated an array of 'smart tools' to support its effective use. CI has been used in many countries including the US, Canada, Europe and Israel. Recently, development in the use of CI in mathematics has been led by Boaler in the UK and the US (Boaler, 2008; Boaler & Staples, 2008).

This kind of R & D is an international resource that provides a valuable foundation and touchstone for education systems elsewhere. Success in implementation elsewhere requires a similar R & D approach to ensure deep understanding and to trial, adapt and further develop effectiveness through tried and tested approaches that are responsive and effective in local contexts.

Policy scepticism and the new opportunity for educational improvement

That principles about effective professional development derived from syntheses of evidence linked to student outcomes are now emerging has particular significance for policy-makers (Timperley & Alton-Lee, 2008). For some decades, there has been somewhat understandable policy scepticism about the efficacy of investment in professional development for teachers. Even given compelling new evidence about the substantive of economic benefit to national GDP of improving the quality of teaching, Hanushek (2005) considered the available evidence base at that time did not justify policy confidence in investment in professional development to improve quality:

While some in-service training and development programs have had success, in general they have been disappointing. Moreover, existing evidence on in-service programs gives insufficient means for selecting a program that is likely to yield significant gains in teaching performance. (Hanushek, 2005, p. 19)

After reviewing a range of survey studies, Chung Wei, Andree and Darling-Hammond (2009) concluded that 'well-designed opportunities are not representative of most teachers' professional development experiences' in the United States. Sparks (2004) drew attention to the professional scepticism of teachers forced to participate in poor quality professional development: 'for far too many teachers in the United States, staff development is a demeaning, mind-numbing experience' (p. 247). He compellingly illustrated the point with the observation of one teacher 'I hope I die during an in-service session, because the transition between life and death would be so subtle' (p. 247). In the *Handbook of research on improving stu-*

dent achievement Sparks (2004) proposed a new agenda for teacher professional development focused on improving student learning. Timperley et al. (2007) found that the effectiveness of professional development was crucial to teacher motivation. When teachers participated in professional development that resulted in substantive improvements for their students, even when their participation was mandatory rather than voluntary, teachers valued the professional development.

Coburn's (2003) *Rethinking scale: Moving beyond numbers to deep and lasting change* provided a salutary caution, given the challenge of translating highly effective professional development into large scale provision. In the light of the failure of so many reforms to lead to lasting improvements in classroom practice, Coburn identified four dimensions of scaling up that need attention: depth, sustainability, spread and shifting reform ownership to schools. Others have extended Coburn's framework to remind policy-makers that, whenever scaling up a professional development reform, it is important to keep goals, purposes and the use of evidence to the fore so as not to lose the focus on improvements for students; and that there needs to be an explicit leadership strategy that attends to structural and organisational changes necessary for deep pedagogical change to occur (Bishop & Berryman, 2010; Cordingley & Bell, 2007).

Within the last five years there has been a growing shift in policy attention as evidence emerges of the potential of high impact professional development to support system lifts. This policy shift is gaining momentum as new evidence of effectiveness becomes available. There are available, for example, a series of reviews of evidence about effective continuing professional development (CPD) carried out in the UK Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) (for example, Cordingley, Bell, Thomason, & Firth, 2005), the New Zealand BES, and new evidence about systemic improvements arising from a focus on teacher learning in jurisdictions such as Ontario, Singapore and Finland (Chung Wei et al., 2009; Levin, 2008).

In 2009 Chung Wei et al. reported 'all around the world, nations seeking to improve their education systems are investing in teacher learning as a major engine for academic success' (p. 28). A 2010 McKinsey report *How the world's most improved school systems keep getting better* reinforced the message. The McKinsey analysis identified 'building the instructional skills of teachers and management skills of principals' (p. 3) as the first of six findings of interventions critical to systemic improvement across 20 nations no matter what the current level of system performance (Mourshed, Chijioke, & Barber, 2010).

Improvement through knowledge building and inquiry cycles

The increasing policy focus on professional development as a critical lever for systemic improvement requires careful attention to evidence about the necessary and sufficient conditions under which teacher professional learning and development translates into improved outcomes for students in a local context. Ascertaining what works in any particular context requires an inquiry approach that builds upon the knowledge available but is locally responsive.

Every context gives rise to its own new challenges. The evidence indicates that it is an inquiry rather than a prescriptive approach that will optimise the effectiveness of professional learning. As Timperley et al. (2007) explain, it is theory, not prescription that travels. In her summary of the BES findings, Timperley (2008)

provides a model of a teacher inquiry and knowledge-building cycle approach that promotes continuous improvement and teacher self-regulation of professional learning. Halbert, Kaser and Koehn (2011) report that teacher use of the professional inquiry diagram with its focus on valued student outcomes, has resulted in higher student gains within their evolving *spiral of inquiry* approach across a large and established network of schools in British Columbia.

In Figure 3, the professional inquiry model has been developed further in the light of other BES findings about the potential for more equitable practice when educationally powerful connections are made with students' lives, identities and families.

Given the potential pitfalls of large scale implementation, such an inquiry approach is needed at every level of change, from school to policy, to ensure ongoing improvement and systemic effectiveness. A cumulative R & D approach is

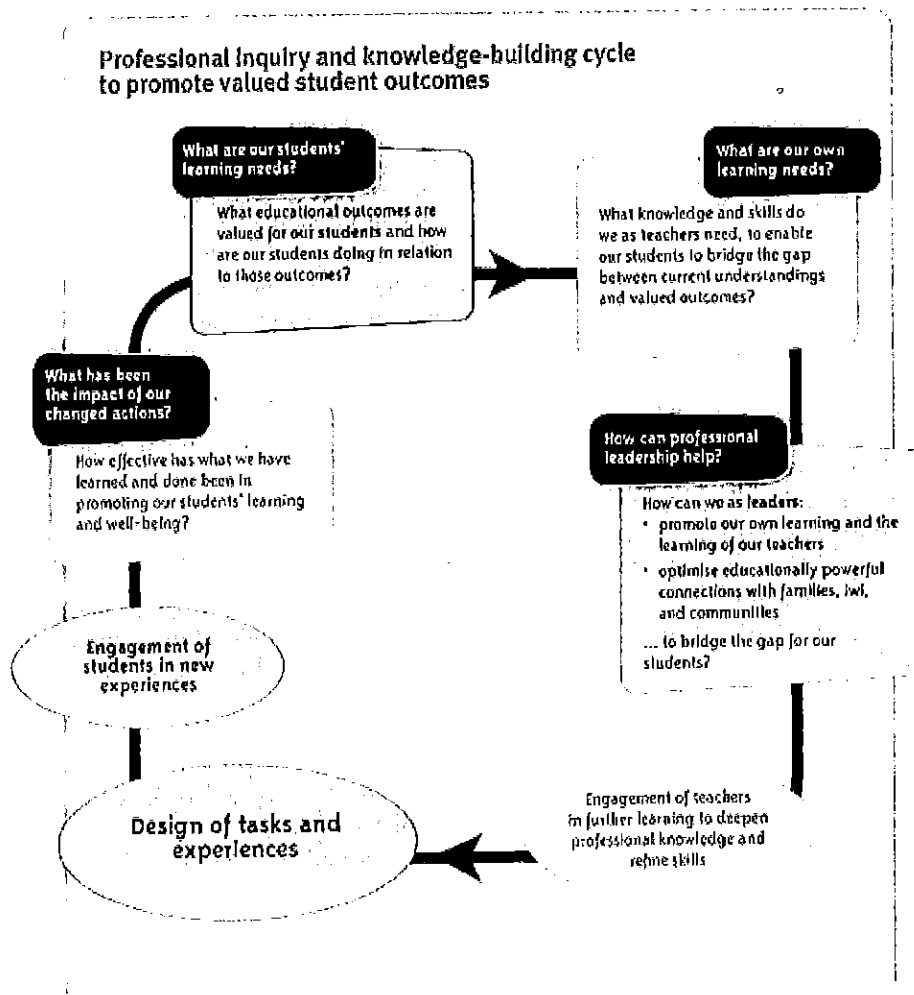


Figure 3. Teacher inquiry and knowledge-building cycles to promote valued student outcomes.

inherently a cyclical inquiry process that builds on knowledge of what makes the most difference at the same time as it fosters the responsiveness and adaptive expertise required for ongoing refinement. Taking a sustained inquiry approach to systemic change at a national level is a challenge for policy because of inherent tensions and project discontinuities across political cycles.

Using evidence for systemic improvement: need and opportunity

Policy workers need to be using inquiry to answer the improvement questions in Figure 3. What education outcomes are valued for our students and how are our students doing in relation to those outcomes at the national level? How effective has what we have learned and done been in promoting all students' learning and well being at a national level?

Mathematics is a particular challenge for New Zealand primary schooling. It is a curriculum area that has been a government policy priority for more than a decade. In contrast to a relatively high performance at secondary level where the specialist knowledge of mathematics teachers is high, average performance at middle primary level remains below the international mean in the Trends in Mathematics and Science Studies (TIMSS) (Caygill & Kirkham, 2008; Mullis, Martin, & Foy, 2008).

The New Zealand Numeracy Development Project (NDP), progressively scaled-up over the last decade, has been a national professional development provision focused on classroom practice. It reached 97% of schools within a self-managing school system. The NDP oriented teachers to focus on student thinking and strategies using diagnostic tools to track learning. The project significantly influenced classroom practice and lifted student achievement in mathematics in English and Māori medium education. The effect sizes for gains on addition and subtraction at primary level, for example, calculated on a national sample across 2003, 2005 and 2007, were well in excess of $d=0.40$ (Young-Loveridge, 2010). Students whose teachers participated in the NDP achieved significantly more highly in Trends in International Mathematics and Sciences Studies than those whose teachers had not (Caygill & Kirkham, 2008).

However, despite some reduction in mathematics achievement disparities in New Zealand from 1994 to 2002, the momentum for positive change weakened and the disparities worsened for Māori and Pasifika primary students in good economic times between 2002 and 2006 (see Table 3). There was also a significant increase in the proportion of Māori students (from 18% to 25%) and Pasifika students (from 23% to 38%) not reaching the low international benchmark between 2002 and 2006. The National Education Monitoring Project found disparities persisted for Māori and worsened for Pasifika over the 2005 to 2009 period (Crooks, Smith, & Flockton, 2010). These monitoring alerts indicate that much more is needed to achieve a systemic lift and greater responsiveness to all New Zealand students.

The Numeracy Development Project brought the challenge of connecting teacher professional learning and student outcomes into sharp focus. An early evaluation revealed that *who* facilitated the professional learning was critical (Higgins, 2005). Facilitators who had the expertise to develop teachers' knowledge and understanding so that they could take an inquiry approach to developing contextually responsive practice were most successful. Those who used a more prescriptive approach created a 'design adherence' mindset in teachers that did not equip them with the

Table 3. 1994/5–2006/7 TIMSS mean trend results for New Zealand Year 5 mathematics achievement for all students and for Māori and Pasifika students. (Source: Caygill & Kirkham, 2008).

Year	International Scaled mean	Mean Score (Standard error) New Zealand students	Range from 5 th to 95 th percentile	Mean Score (Standard error) Māori students	Mean Score (Standard error) Pasifika students
2006	500	492 (2.3)	284	453 (4.4)	427 (5.1)
2002	500	496 (2.1)	273	479 (4.8)	464 (6.3)
1998	Only NZ	481 (5.6)	312	445 (7.3)	416 (15.1)
1994	500	469 (4.4)	316	427 (8.2)	412 (11.0)

adaptive expertise necessary for greater responsiveness to and success with student learning. A further weakness of the 'design adherence' approach was that it did not support teachers to develop the self-regulatory skills that would enable them to use assessment for the purposes of professional inquiry and ongoing improvement.

Successive evaluations indicated that there was an issue with the sufficiency of the professional development opportunities. For example, where scaling up was attempted too quickly, there was insufficient time to develop capability in facilitators, and emotional issues around change were not adequately addressed. Under such conditions teachers did not receive responsive, knowledgeable support, sufficient access to the knowledge they needed or the multiple opportunities needed to learn and apply the new information. Wider spread across the schooling system was achieved but at the cost of depth and sufficiency needed for deeper and sustainable change (these findings are apparent in evaluations of the New Zealand Numeracy Development Project, including Young-Loveridge (2010), which are all available online at <http://nz.maths.co.nz/annual-research-and-evaluation-reports-and-compendium-papers>).

Successive evaluations also revealed the importance of the proactive involvement of school leadership in creating effective conditions for the professional development and, the critical role of lead teachers in the success or otherwise of organisational change. For example, when lead teachers of numeracy also had positional authority in the school, organisational change was more successful (Higgins & Bonne, 2010).

Using and building on the knowledge that has been garnered to date about what did and did not work in the NDP will be critical for lifting achievement in mathematics for New Zealand primary students. But it will not be enough. New evidence from the best evidence synthesis series including that focused on effective mathematics teaching (Anthony & Walshaw, 2007), and the new international literature on highly effective educational reform (Levin, 2008) provide valuable resources and opportunities for informing a systemic lift in achievement, and in the New Zealand context, accelerating progress for Māori and Pasifika students, low achievers and those disadvantaged by socio-economic inequality.

Using what we know about what works and what makes a bigger difference in education

Systemic improvement calls for increased knowledge, understanding and use of teaching strategies that make a bigger difference to a range of valued outcomes in education. Teachers need viable alternatives to practices that do harm or are less effective. To be optimally effective, professional development needs to build upon the findings of research and development about professional knowledge and skills that have a greater positive impact on student outcomes for time invested. While the value of this approach to effectiveness may seem self-evident, Timperley (2008) observes that 'unproven ideas continue to sweep through educational jurisdictions' (p. 10). In New Zealand, strategies that research evidence has revealed to be potentially detrimental to student outcomes are reported to be commonly used with our lowest achievers. Even research-based approaches can be harmful if the underpinning research has not attended to impacts on student outcomes. A prominent example in New Zealand is the prevalent use of learning styles matching approaches that can ghettoise Māori and Pasifika students into kinaesthetic activities

with concrete materials and procedural activity while other students engage in meta-cognitive strategy instruction (Alton-Lee, 2003; Higgins, 2001). Other potentially harmful activities are fixed ability grouping, the allocation of the least trained adults (teacher aides) to work with the lowest achieving students, and public ranking and labelling of students by achievement levels on classroom walls (Hattie, 2009).

Parental help with mathematics homework (as opposed to parental support for their children during their homework activity) has been found to have an ongoing negative impact on student achievement (Wylie, Thompson, & Lythe, 2001). These negative effects highlight the importance of high impact R & D informed interventions to assist parents in supporting their children's mathematics homework. For example, the development and use of mathematics games libraries to foster positive parent-child interaction that builds children's capabilities and confidence.

In their summary of 10 BES findings about effective teaching practices in mathematics Anthony and Walshaw (2009) first drew attention to the importance of building a caring mathematics learning community. While carefully identifying the value of both teacher-directed and individual student work, they highlighted as a priority the strengthening of student capability to cooperate and collaborate in their learning.

Creating a caring, classroom learning communities is a challenge in many countries but particularly so for New Zealand. On a self-report index that included items such as being made fun of, called names, being excluded, being hurt and having something stolen, New Zealand was found to have the second lowest rating for student safety in the peer culture of the 35 participating countries in the Trends in International Mathematics and Science Study (Mullis et al., 2008). A decade earlier a TIMSS international comparison found New Zealand primary students' fear of being hurt by peers to be amongst the highest reported (Garden, 1997, p. 168). A new report from the Prime Minister's Chief Science Advisor concludes that 'adolescents in New Zealand relative to those in other developed countries have a high rate of social morbidity' (Gluckman, 2011a, p. 1).

For value for investment from a policy perspective what is needed is to develop capability in high impact teaching approaches that advance academic and social outcomes for all students while they also reduce disparities. For teachers and principals, also, this means working smarter rather than harder for educational improvement and intensifying peer supports for student learning.

New high impact research and development building on BES findings

Dr Roberta Hunter, a leading Pasifika mathematics education researcher in New Zealand, developed a high impact intervention in mathematics teaching, drawing on a range of high impact R & D including *Complex Instruction*. Hunter (2007, 2008) designed a year-long school-based collaborative professional development programme to assist teachers to strengthen student co-operation skills and to build a community of mathematical inquiry in their mixed age primary classes. This intervention was part of a wider project organised by the writers of the *Effective pedagogy in mathematics BES* co-directors of the Centre for Research in Mathematics Education at Massey University.

The intervention brought about significant changes in teacher knowledge and pedagogy and in student behaviour and mathematical practices. These changes were achieved with predominantly Māori and Pasifika students in a low socio-economic

status school. The effect sizes for the gains in the two classes that were the focus of the research were very large: $d=2.39$ and $d=2.53$ (Alton-Lee, Hunter, Pulegatoa-Diggins, & Sinnema, 2011). That is the equivalent of several years' progress in just one year. The researcher traced the process of teacher change for both teachers who had participated previously in the Numeracy Development Project without receiving sufficient support to bring about substantive improvements for their students that the new professional learning enabled.

The teachers worked to develop classroom learning communities in which students learned to engage with the teacher and each other in mathematical inquiry, reasoning and argumentation. The teachers focused on creating a safe peer environment where students were able to take the intellectual risks needed to accelerate their learning. By terms 3 and 4, the students' comments indicated a strongly functioning student learning community. For example, Wiremu, aged 10, told another boy in his group to stop his disrespectful behaviour towards a female peer so she could take a risk in her public participation in the mathematical discourse: 'Don't dis her, man, when she is taking a risk' (Hunter, 2007, p. 111).

Within the Iterative Best Evidence Synthesis Programme we have consulted with teachers and principals to get advice about developing *BES Cases* that describe and explain the changes in teaching practices in such high impact interventions. The *BES Cases* illuminate also the professional learning and development that enabled teachers to improve their practice and professional leadership supports that made the changes possible. This new form of *BES Cases* is a practical tool to explain the *how* of the findings of the best evidence syntheses in a way that is helpful for teachers. The *BES Cases* under development each includes the *Professional Inquiry and Knowledge Building Cycle* diagram and a set of specific questions for adaptive use in other school contexts (Alton-Lee et al., 2011). Both primary and secondary teacher unions have given feedback to the Iterative BES Programme that student group work is a weakness in practice and have advocated for the provision of professional development to support an evidence-based approach to cooperative group work in New Zealand. For systemic improvement to occur, there would need to be policy commitment to creating the conditions for effective scaling up of professional development, and a strategy to build the necessary expertise.

Knowledge mobilisation: a caution

The potentially large benefits of effective student group work on academic and social outcomes have long been documented in educational research literature (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Slavin, 1980; Hattie, 2009) but the use of an evidence-based approach to cooperative group work in schooling remains a challenge.

Two decades ago, Galton and Williamson (1992) in *Group work in the primary classroom* illuminated the implications of this research for teachers and drew attention to the problem that the theoretical acceptance of the power of group work in educational research was not matched by its use in practice. In the frontispiece to their book they pointed out that group work in practice 'can mean almost anything from group seating, as a technique of classroom management, to full collaborative learning'. Their account of successive cycles of research and development in the UK and Australia grappling with the difficulties of beginning, monitoring and maintaining effective group work in primary classrooms was invaluable in informing my

own work in teacher education in the early 1990s. For example, their *Classification of Group Processes* into authority, negotiating, supportive and non-cooperative roles was a useful tool for teachers and students to reflect upon and use to strengthen student interaction skills (p. 161). Galton and Williamson's (1992) landmark book also foreshadowed the challenges for professional learning that are the focus of this article.

After a continuing contribution to knowledge about effective cooperative group work at both primary and secondary level (for example: Galton, 2009), Galton and Hargreaves (2009) again highlighted the persistent neglect of group work in educational practice 'although... there are many studies that attest to the value of group working as effective pedagogy, both in terms of improving pupil attainment and also attitudes, particularly racial ones, the fact remains that in many classrooms groupwork is still a neglected art' (p. 1). In focusing an entire volume of the CJE on findings about effective group work across a range of countries, Galton and Hargreaves sought that the special edition would stimulate more structured group work in classrooms around the world, and further research that builds upon the principles of effectiveness that have emerged in this research field.

Slavin (2010) also commented recently on the historical failure to translate such compelling evidence of effectiveness into practice.

In comparison with schooling practices that are often supported by governments – such as tutoring, technology use and school restructuring – co-operative learning is relatively inexpensive and easily adopted. Yet, thirty years after much of the foundational research was completed, it remains at the edge of school policy. This does not have to remain the case: as governments come to support the larger concept of evidence-based reform, the strong evidence base for co-operative learning may lead to a greater focus on this set of approaches at the core of instructional practice. In the learning environments of the twenty-first century, co-operative learning should play a central role. (Salvin, 2010, p. 174)

Why does this approach remain at the edge of school policy, given such a long-standing and compelling evidence base about a teaching approach that when well-implemented, demonstrably lifts achievement for low and high achievers, advances students' self-regulatory and social skills, reduces bullying and racism and creates more equitable learning environments? The compelling question about evidence for educational improvement is why there is so little progress in take-up at a systemic level in many jurisdictions.

Cooper, Levin and Campbell (2009) make the case for a new research field to develop evidence about effective knowledge mobilisation of educational evidence in policy and practice. In considering barriers to the use of evidence, Pawson (2002) proposed that research evidence is 'softly spoken' because 'empirical inquiry simply cannot make its voice heard amidst the clatter of other, political imperatives on policy-making' (p. 227). In an overview of research about the use of evidence in policy, Moore (2006) found that a competing plethora of ideas are presented to policy-makers as if they were the result of reliable research. Amongst other findings she reported that the use of evidence in policy depends upon high level leadership commitment, ownership of evidence and resourced processes that are integrated into existing organisational systems and ways of working. In recognition of such challenges the Chief Science Advisor to the New Zealand Prime Minister has recently

published a discussion paper: *Towards better use of evidence in policy formation* (Gluckman, 2011b).

Educational R & D: a policy resource for making a bigger difference

Communities are seeking better academic and social outcomes from schooling for all students. This is a fundamental shift in the function of schooling that will not be advanced through magical and/or 'teachers must try harder' policy discourses or a rediscovering the wheel mentality.

In their consideration of how the world's most improved schooling systems keep getting better Mourshed, Chijioke and Barber (2010) conclude that schooling systems can make significant gains in six years or less. Their second major finding was that there is too little focus on 'process' in the improvement debate. This article illustrates how sustained cycles of inquiry-driven high-impact educational research and development can be critical to the 'what', 'why' and 'how' of systemic improvement. When high-impact R & D informs effective professional learning, we have powerful evidence of demonstrable and even dramatic improvements both academic and social outcomes for diverse, high and low achievers over multiple settings.

Internationally, R & D in education has not generally been a priority. An OECD (2003) report *Knowledge management: New challenges for educational research* identified the relatively low proportion of funding afforded educational R & D, and the challenges this raises for knowledge societies:

A rough estimate of the level of educational R & D as a percentage of total expenditure on education is on average less than 0.3% in six countries for which data are available. This is a very small figure when education is compared with other knowledge sectors, for example, the health sector where between 5–10% of the total health expenditure in public and private sectors is directed to R & D. (OECD, 2003, p. 11)

In an OECD (2007, p. 3) report the Director of OECD's Centre for Educational Research and Innovation argued the case that:

The issues of effective relationships between research and policy-makers, capacity building within those domains, and the importance of allocating scarce resources in the most efficacious manner remain as important as they were 10 years ago.

(Using) evidence for educational improvement

The evidence presented in this article demonstrates that effective professional development is a major policy lever for systemic improvement in education, and highlights how R & D can make an important contribution by identifying practices that have significant positive impact, supporting the development of effective professional learning and enabling ongoing improvement. Systemic improvement requires policy, research and professional leadership to prioritise and create the conditions for productive professional learning.

The message for both policy, research and practice – even more so in straightened economic times – is that we should build on what we have learned about the most effective ways of making a difference for the learning and well-being of our

diverse students. And that we should invest strategically in continuing high impact R & D so that we can avoid policy churn, ensure the disciplined development of productive educational innovation and grow capability.

The best evidence synthesis series provides trustworthy information to guide educational policy and practice about the conditions, both necessary and sufficient, for professional development that has transformative impact at system level. This knowledge resource shows the potential for improvement when bodies of evidence (such as the lessons from Maurice Galton's seminal work on cooperative learning) are acted on. Conversely, it highlights the loss of opportunities to advance excellence, equity and social cohesion through education when high impact R & D is ignored and effective professional learning neglected.

For those of us who work in and for education, the challenge is to use the evidence.

Note

An early version of this article was presented to the General Assembly of Fellows of the International Academy of Education, Limassol, Cyprus September 2008.

Note on contributor

Adrienne Alton-Lee is the Chief Education Advisor who leads the New Zealand Ministry of Education's Iterative Best Evidence Synthesis (BES) Programme. Her role is to strengthen the development and use of the evidence-base informing policy and practice in education to support systemic improvement for diverse (all) learners. She is a Fellow of the International Academy of Education. She was formerly a teacher, classroom researcher, professor and an associate editor of *Teaching and Teacher Education*. She has published in leading educational journals including the *Harvard Educational Review*, the *Elementary School Journal*, the *International Journal of Inclusive Education* and *Review of Research in Education*; and is the author of the Ministry of Education's first BES: *Quality teaching for diverse students in schooling best evidence synthesis iteration* (<http://www.educationcounts.govt.nz/goto/BES>).

References

- Ainscow, M., Chapman, C., Dyson, A., Gunter, H., Hall, D., Kerr, K., McNamara, O., Muijs, D., Raffo, C., & West, M. (2010). Social inequality: Can schools narrow the gap? Insight 2, British Educational Research Association. Retrieved August 29, 2011, from <http://www.bera.ac.uk/files/2011/01/Insight2-web.pdf>
- Aitken, G., & Sinnema, S. (2008). *Effective pedagogy in social sciences Tikanga ā iwi: Best evidence synthesis iteration*. Wellington: Ministry of Education. Retrieved August 29, 2011, from <http://www.educationcounts.govt.nz/publications/series/2515/32879/35263>
- Alton-Lee, A. (2003). *Quality teaching for diverse students in schooling: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.
- Alton-Lee, A. (2004). Improving education policy and practice through an Iterative Best Evidence Synthesis Programme. Invited address to OECD-US Seminar, Evidenced-Based Policy Research, Washington DC, April 19–20. Retrieved August 29, 2011, from <http://www.educationcounts.govt.nz/topics/BES/draft-case-studies>.
- Alton-Lee, A., Hunter, R., Pulegatoa-Diggins, C., & Sinnema, C. (2011). *Case 1: Developing communities of mathematical inquiry. Quality teaching for diverse (all) learners in schooling. Second iteration best evidence synthesis in progress*. Retrieved from <http://www.educationcounts.govt.nz/goto/BES>
- Alton-Lee, A., & Nuthall, G., with Patrick, J. (1995). Reframing classroom research: A lesson from the private world of children. In G. Capella Noya, K. Geismar, & G. Nicoleau (Eds.). *Shifting histories: Transforming education for social change*. Cambridge, MA: Harvard Educational Review. Reprint series no. 26.

- Anthony, G., & Walshaw, M. (2007). *Effective pedagogy in mathematics/pāngarau: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.
- Anthony, G., & Walshaw, M. (2009). *Effective pedagogy in mathematics: Educational practices series, 19*. Retrieved August 29, 2011, from <http://www.ibe.unesco.org/en/services/publications/educational-practices.html>
- Biddulph, F., Biddulph, J., & Biddulph, C. (2003). *The complexity of community and family influences on children's achievement in New Zealand: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.
- Bishop, R., & Berryman, M. (2010). *Scaling up education reform: Addressing the politics of disparity*. Wellington: New Zealand Council for Educational Research (NZCER) Press.
- Bishop, R., & Glynn, T. (1999). *Culture counts: Changing power relations in education*. Palmerston North: Dunmore Press.
- Boaler, J. (2008). Promoting 'relational equity' and high mathematics achievement through an innovative mixed ability approach. *British Educational Research Journal*, 34(2), 167–194.
- Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers' College Record*, 110(3), 608–645.
- Bossert, S. (1979). *Tasks and social relationships in classrooms: A study of instructional organisation and its consequences*. London: Cambridge University Press.
- Brandon, P.R., Taum, A.K.H., Ayala, C.C., Young, D.B., Gray, M.E., Speitel, T.W., Nguyen, T.T.T., & Pottenger, F.M. III. (2007). *Phase-1 Study of the effects of professional development and long-term support on program implementation and scaling up*. Final report. Retrieved August 29, 2011, from http://www.hawaii.edu/crdg/programs/documents/evaluation/executive_summary.pdf
- Caygill, R., & Kirkham, S. (2008). *Mathematics. Trends in year 5 mathematics achievement 1994 to 2006. New Zealand results from three cycles of the Trends in International Mathematics and Science Study (TIMSS)*. Wellington, New Zealand: Ministry of Education.
- Chung Wei, R., Andree, A., & Darling-Hammond, L. (2009). How nations invest in teachers: High achieving nations treat their teachers as professionals. *Educational Leadership*, 66(5), 28–33.
- Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3–12.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, E., Bianchini, J., Cossey, R., Holthuis, N., Morphew, C., & Whitcomb, J. (1997). What did students learn? 1982–1994. In E. Cohen & R. Lotan (Eds.), *Working for equity in heterogeneous classrooms: Sociological theory in practice* (pp. 137–165). New York: Teachers College Press.
- Cohen, E., Lotan, R., & Hothuis, N. (1997). Organising the classroom for learning. In E. Cohen & R. Lotan (Eds.), *Working for equity in heterogeneous classrooms: Sociological theory in practice* (pp. 31–43). New York: Teachers College Press.
- Cole, M. (1998). Racism, reconstructed multiculturalism and antiracist education. *Cambridge Journal of Education*, 28(1), 37–48.
- Cooper, A., Levin, B., & Campbell, C. (2009). The growing (but still limited) importance of evidence in education policy and practice. *Journal of Educational Change*, 10(2–3), 159–171.
- Cordingley, P., & Bell, M. (2007). Transferring learning and taking innovation to scale. Centre for the Use of Research Evidence in Education. Thinkpiece prepared for The Innovation Unit. Retrieved August 29, 2011, from <http://www.innovationunit.org/resources/transferring-learning-and-taking-innovation-scale>
- Cordingley, P., Bell, M., Thomason, S., & Firth, A. (2005). *The impact of collaborative continuing professional development (CPD) on classroom teaching and learning. Review: How do collaborative and sustained CPD and sustained but not collaborative CPD affect teaching and learning?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Crooks, T., Smith, J., & Flockton, L. (2010). *Mathematics assessment results 2009*. National Education Monitoring Report 52. Dunedin: Educational Assessment Research Unit. Retrieved August 29, 2011, from <http://nemp.otago.ac.nz>

- Dahl, R.F. (1997). Organizational factors and the continuation of a complex instructional technology. In E.G. Cohen & R.A. Lotan (Eds.), *Working for equity in heterogeneous classrooms – sociology theory in practice* (pp. 260–274). New York: Teachers College Press.
- Davidson, J. (2011). Comment, BES introductory flyer. Wellington: Ministry of Education. Retrieved August 29, 2011, from <http://www.educationcounts.govt.nz/goto/BES>
- Donn, M., & Schick, R. (1995). *Promoting positive race relations in New Zealand schools: Me mahi tahi tatou*. Wellington: Research Division, Ministry of Education.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53, 159–199.
- Filby, N. (1997). A viewpoint on dissemination. In E. Cohen & R. Lotan (Eds.), *Working for equity in heterogeneous classrooms: Sociological theory in practice* (pp. 277–286). New York: Teachers College Press.
- Galton, M. (2009). Working together – learning together: Cooperative working in the classroom. In M. Galton, S. Steward, L. Hargreaves, C. Page, & A. Pell (Eds.), *Motivating your secondary class* (pp. 119–140). London: Sage.
- Galton, M., & Hargreaves, L. (2009). Group work: still a neglected art? *Cambridge Journal of Education*, 39(1), 1–6.
- Galton, M., & Williamson, J. (1992). *Groupwork in the primary classroom*. London: Routledge.
- Garden, R. (1997). *Mathematics and science performance in middle primary school: Results from New Zealand's participation in the third mathematics and science study*. Wellington: Ministry of Education.
- Gluckman, P. (2011a). Improving the transition: Reducing social and psychological morbidity during adolescence. A report from the Prime Minister's Chief Science Advisor, Auckland. Retrieved August 29, 2011, from <http://www.pmcsa.org.nz>
- Gluckman, P. (2011b). *Towards better use of evidence in policy formation: A discussion paper*. Auckland: Office of the Prime Minister's Science Advisory Committee. Retrieved August 29, 2011, from <http://www.pmcsa.org.nz/wp-content/uploads/2011/04/Towards-better-use-of-evidence-in-policy-formation.pdf>
- Halbert, J., Kaser, L., & Koehn, D. (2011, January). *Spirals of inquiry: Building professional inquiry to foster student learning*. Paper presented at the annual meeting of the International Congress of School Effectiveness and Improvement, Limassol, Cyprus.
- Hanushek, E. (2005). *Economic outcomes and school quality*. International Academy of Education and The International Institute for Educational Planning, UNESCO. Retrieved August 29, 2011, from <http://unesdoc.unesco.org/images/0014/001410/141027e.pdf>
- Harker, R. (2007). *Ethnicity and school achievement in New Zealand: Some data to supplement the Biddulph et al. (2003) Best Evidence Synthesis*. Unpublished report to the Iterative Best Evidence Synthesis Programme. Wellington: Ministry of Education.
- Hattie, J. (2009). *Visible teaching – visible learning: A synthesis of 800+ meta-analyses on achievement*. London: Routledge.
- Higgins, J. (2001). *Developing numeracy: Understanding place value*. Report to the Ministry of Education. Wellington: Ministry of Education.
- Higgins, J. (2005). Pedagogy of facilitation: How do we best help teachers of mathematics with new practices? In H.L. Chick & J.L. Vincent (Eds.), *Proceedings of the 29th annual conference of the International Group for the Psychology of Mathematics Education*, 3 (pp. 137–144). Melbourne: PME.
- Higgins, J., & Bonne, L. (2010). The role of the numeracy lead teacher in promoting the goals of the Numeracy Development Projects. In R. Averill, B. Bicknell, L. Bonne, T. Burgess, N. Hāwera, R. Hunter, M. Johnston, C. Linsell, G. Lomas, J. Mills, P. Nicholas, B. Stevenson, A. Tagg, M. Taylor, P. Te Maro, G. Thomas, T. Trinick, J. Ward, & J.M. Young-Loveridge (Eds.), *Findings from the Numeracy Development Project 2009* (pp. 129–138). Wellington: Ministry of Education. Retrieved August 29, 2011, from <http://www.nzmaths.co.nz/node/7516>
- Hunter, R. (2007). *Teachers developing communities of mathematical inquiry*. Albany, Massey University: Unpublished doctoral thesis. Retrieved August 29, 2011, from <http://www.educationcounts.govt.nz/topics/BES/nz-education-theses-database>

- Hunter, R. (2008). Facilitating communities of mathematical inquiry. In M. Goos, R. Brown, & K. Makar (Eds.), *Navigating currents and charting directions* (Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, Vol. 1, pp. 31–39). Brisbane: MERGA.
- Johnson, D.W., Maruyama, G., Johnson, R.T., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta analysis. *Psychological Bulletin*, 89, 47–62.
- Levin, B. (2008). *How to change 5000 schools: A practical and positive approach for leading change at every level*. Cambridge, MA: Harvard Education Press.
- Levin, B. (2009). Enduring issues in urban education. *Journal of Comparative Policy Analysis: Research and Practice*, 11(2), 181–195.
- Levin, B., & Fullan, M. (2009). International foreword. In V. Robinson, M. Hohepa, & C. Lloyd, *School leadership and student outcomes: Identifying what works and why. Best evidence synthesis iteration* (pp. 13–15). Wellington: Ministry of Education. Retrieved August 29, 2011, from <http://www.educationcounts.govt.nz/publications/series/2515/60169/60170>
- Lipman, P. (1997). Restructuring in context: A case study of teacher participation and the dynamics of ideology, race and power. *American Educational Research Journal*, 34(1), 3–37.
- Moore, P. (2006). *Evidence based policy report*. Iterative Best Evidence Synthesis Programme Evidence Based Policy Project Report, Wellington, New Zealand. Retrieved August 29, 2011, from www.educationcounts.govt.nz/publications/series/2515/27547/27543
- Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. Retrieved August 29, 2011, from http://www.mckinsey.com/client-service/Social_Sector/our_practices/Education/Knowledge_Highlights/How%20School%20Systems%20Get%20Better.aspx
- Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Neves, H.A. (2007). The relationship of talk and status to second language acquisition of young children. In E. Cohen & R. Lotan (Eds.), *Working for equity in heterogeneous classrooms: Sociological theory in practice* (pp. 181–192). New York: Teachers College Press.
- Nuthall, G. (1999). Learning how to learn: The evolution of students' minds through the social processes and culture of the classroom. *International Journal of Educational Research*, 31(3), 141–256.
- OECD. (1995). *Educational research and development: Trends issues and challenges*. Paris: OECD.
- OECD. (2003). *Knowledge management: New challenges for educational research*. Paris: OECD.
- OECD. (2007). *Knowledge management: Evidence in education linking research and policy*. Paris: CERI, OECD. Retrieved August 29, 2011, from <http://www.sourceoecd.org/education/9789264033665>
- OECD. (2010). PISA 2009 at a glance. Retrieved August 29, 2011, from <http://www.oecd.org/dataoecd/31/28/46660259.pdf>
- Osler, A., & Starkey, H. (1999). Rights, identities and inclusion: European action programmes as political education. *Oxford Review of Education*, 25(1/2), 199–215.
- Pawson, R. (2002). Evidence and policy and naming and shaming. *Policy Studies*, 23(3), 211–230.
- Robinson, V., Hohepa, M., & Lloyd C. (2009). *School leadership and student outcomes: Identifying what works and why: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education. Retrieved August 29, 2011, from <http://educationcounts.govt.nz/goto/BES>
- Robinson, V.M.J., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes. An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44(5), 635–674.

- Seixas, P. (2001). Review of research on social studies. In V. Richardson (Ed.), *Handbook of research on teaching*, 4th edition (pp. 545–565). Washington, DC: American Educational Research Association.
- Shaver, J. (1999). Social Studies: Cognitive prejudice reduction. In G. Cawelti (Ed.), *Handbook of research on improving student achievement* (pp. 180–197). Virginia: Educational Research Service.
- Slavin, R. (1980). Co-operative learning. *Review of Educational Research*, 50(2), 315–342.
- Slavin, R. (2010). Co-operative learning: what makes group-work work? In H. Dumont, D. Instance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 161–178). Paris: OECD.
- Snook, I. (2008). 'Tail' of underachievement is tale of poverty. Retrieved August 29, 2011, from <http://www.scoop.co.nz/stories/PO0804/S00419.htm>
- Sparks, D. (2004). Focusing staff development on improving the learning of all students. In G. Cawelti (Ed.), *Handbook of research on improving student achievement* (pp. 198–208). Virginia: Educational Research Service.
- Timperley, H. (2008). *Teacher professional learning and development: Educational practices series-18*. Retrieved August 29, 2011, from <http://www.ibe.unesco.org/en/services/publications/educational-practices.html>
- Timperley, H., & Alton-Lee, A. (2008). Reframing teacher professional learning: An alternative policy approach to strengthening valued outcomes for diverse learners. *Review of Research in Education*, 32(1), 328–369.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education. Retrieved August 29, 2011, from <http://educationcounts.govt.nz/goto/BES>
- University of Hawaii. (2002). Foundational approaches in science teaching (FAST). In J. Killion (1999). *What works in the middle: Results-based staff development*. Oxford, OH: National Staff Development Council. Retrieved August 29, 2011, from <http://www.learningforward.org/midbook/index.cfm>
- Wylie, C., Thompson, J., & Lythe, C. (2001). *Competent children at10: Families, early education, and schools*. Wellington, New Zealand: New Zealand Council for Educational Research.
- Yamamoto, K. (2007, March). *Against all odds: Tales of survival of the foundational approaches in science teaching (FAST)* project. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Young, D. (1999). *Foundational approaches in science teaching (FAST)*. (Summary of evaluations). Retrieved August 29, 2011, from <http://www.hawaii.edu/crdg/sections/science/documents/fasteval.pdf>
- Young-Loveridge, J. (2010). A decade of reform in mathematics education: Results for 2009 and earlier years. In R. Averill, B. Bicknell, L. Bonne, T. Burgess, N. Hāwera, R. Hunter, M. Johnston, C. Linsell, G. Lomas, J. Mills, P. Nicholas, B. Stevenson, A. Tagg, M. Taylor, P. Te Maro, G. Thomas, T. Trinick, J. Ward, & J.M. Young-Loveridge (Eds.), *Findings from the New Zealand Numeracy Development Projects 2009*. Wellington: Ministry of Education. Retrieved August 29, 2011, from <http://nzmaths.co.nz/annual-research-and-evaluation-reports-and-compendium-papers>