Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author’s right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. [http://researchspace.auckland.ac.nz/feedback](http://researchspace.auckland.ac.nz/feedback)

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.
A Theory for Schooling Improvement:
Consistency and Connectivity to Improve Instructional Practice

Brian Annan

ABSTRACT

This thesis investigates the problem of how to speed up the process through which professional educators learn how to significantly improve disadvantaged students' academic achievement. The problem is addressed through three questions: (i) What are the most effective national and international examples of school improvement? (ii) What is the condition of the evidence base for making claims of effectiveness? (iii) What can be learned about developing and implementing effective school improvement from those national and international examples? The thesis begins by searching international and national school improvement literature to find those initiatives with the strongest evidence of effectiveness. One initiative in England (the National Literacy and Numeracy Strategies) and four initiatives in the United States (Success For All, Direct Instruction, The School Development Programme & a district-wide reform in New York District #2) were considered to have strong evidence of effectiveness. Two initiatives in New Zealand (the Numeracy Development Project & the Strengthening Education in Mangere and Otara project commonly called SEMO) had evidence that showed promise. It is argued that patterns of investment in different types of evaluation and ease of access to achievement information account for the difference between the strong international evidence and promising evidence in New Zealand. A series of investigations in the middle of the thesis focus on the processes set up in the initiatives to help practitioners learn effective reform practices. Three models of learning processes are developed which reveal a strong preference for vertical learning in England and the United States and a more balanced vertical-horizontal learning preference in New Zealand. Despite those contrasts, three characteristics were found to be common to all seven effective initiatives. They are a sharp focus on instructional improvement, a set of standardised practices, and, learning connections to transfer the reform ideas into practice. The latter part of the thesis transforms those three characteristics into a theory for schooling improvement which contribute to a faster and more effective reform process.
ACKNOWLEDGEMENTS

Most accomplishments in the field of school improvement require many minds and this doctoral research is no exception. It is not possible to name all the Ministry and school colleagues, community members, scholars, friends and family members as well as numerous disadvantaged students I taught in my earlier career that either inspired this research or provided feedback on it. I am especially indebted to my past managers in the Ministry, Kathy Smith and Mary Sinclair, and my current managers, Colin McGregor and Rawiri Brell, for approving and supporting the research. Alongside my managers, special thanks goes to my Ministry colleagues Te Kepa Stirling, Elena Fa-amoe-Timoteo, Susan Warren, Ray Webb, Martin Connelly, John Good, Malcolm Hyland and Adrienne-Alton Lee for their invaluable feedback that helped to improve my thinking about systemic change. I also greatly appreciated Raewyn Corner’s outstanding administrative support.

It has also been a privilege to be heavily influenced by the positive world views of many inspirational school and community members who work hard to make their school improvement initiatives successful. In particular, I would like to thank Karen Mose and Corinne Hansell and their colleagues in Mangere who led the way in focusing on raising student achievement and Pat Kake and Colleen Murray and their colleagues who also showed that community governance of school improvement could work. Equally influential were researchers and developers willing to work with and for those groups. Particular thanks go to Professor Stuart McNaughton, Professor Helen Timperley, Dr Gwenneth Phillips, Dr Shelley MacDonald, Dr Mei Lai, Jill Pritchard and Kerry Taylor and her associates Jan Hill and Kay Hawke.

As far as scholars in the field of school improvement go, thanks to my two supervisors from The University of Auckland, Professor Viviane Robinson and Professor Roger Dale for guiding me through this research project. Viviane constantly raised the quality bar with her outstanding conceptual and theoretical challenges and her attention to evidence-informed detail. I want to especially thank Professor Robinson for her patience and perseverance in helping me move on from trying to write a thousand ideas into every paragraph. Professor Dale’s distance supervision from England with the occasional face-to-face debates also tested my
thinking particularly around the complex role of the state in school improvement. He also made sure that the finished product was readable for the diverse professional audience that this thesis is intended for. There are two other scholars from my past that I wish to acknowledge for steering me in the right direction. One is Colin Lankshear, a deep-thinking lecturer from my undergraduate studies who was passionate about high quality education for everyone. The other is Miss Moffat, my sixth form English teacher who turned a sufficient amount of my interest away from the sports field to the rigours of academia for the latter to become a life-long pursuit.

I want to thank family and friends. To my wife, Jean, and my two daughters, Rebecca and Kimberley, goes my heart-felt thanks for giving me the space to think deeply and to tap away on the laptop computer which has been glued to my lap for several years. Also Jean’s clarity of academic thought and positive view of life and the world helped me through some of those really difficult times one has during such a project. To my mum, Lois, and my dad, Dave, I will be forever indebted to them as will my older brother, Wayne, and my two younger sisters, Jennifer and Heather, for teaching us the meaning of resilience early in life and for giving us a work ethic that one needs to have to take on and finish challenging projects like this one. My thanks also go to those friends who listened along the way, in particular Russell Burt and Debbie Ayres. To all my other friends, I can only hope that they are willing for me to visit again after withdrawing into the reclusive mode necessary to complete this project.

Finally, many of my former students from disadvantaged communities also come to mind as inspiration for completing this research and to continue on in the quest to solve the underachievement problem in New Zealand. Tone Kolose and Rewa Morris are two prime examples that beat the odds. They amongst many others filled our classroom at Bader Intermediate in the district of Mangere in the early 1980’s with their character and passion for life. Tone went on to become a school principal and Rewa became a teacher aide. I bump into them from time to time and our eyes light up with the fond memories of learning together two decades ago. We reminisce and laugh and catch up with what we have learnt more recently. Those moments remind me of the reason for our endeavour.
TABLE OF CONTENTS

Abstract...........................................................................................................i
Acknowledgements.........................................................................................ii
Table of Contents............................................................................................iv
List of Figures....................................................................................................ix
List of Tables.....................................................................................................x

PART I: SETTING THE SCENE.........................................................................1

Chapter 1: Introduction......................................................................................5

1.1. The research problem.................................................................................6
1.2. Purpose of the thesis...................................................................................9
1.2.1. Effectiveness........................................................................................10
1.2.2. Development and implementation of schooling improvement initiatives.....14
1.3. Research questions.....................................................................................15
1.4. Methodology...............................................................................................17
1.4.1. Ethics..................................................................................................18
1.4.2. Data gathering and accuracy checks.....................................................19
1.5. Organisation of the thesis..........................................................................24

Chapter 2: Macro Policy Contexts.................................................................31

2.1. The macro policy context in England.......................................................33
2.2. The macro policy context in the United States.........................................36
2.3. The macro policy context in New Zealand...............................................40
2.3.1. Influence of the 1989 administrative reforms........................................41
2.3.2. Influence of the intervention framework..............................................44
2.4. Summary and conclusions........................................................................48
Chapter 6: Learning Models

6.1. The learning process model for England's literacy and numeracy strategies

6.1.1. Learning contexts

6.1.2. Vertical and horizontal learning

6.1.3. Barrier management

6.1.4. Summary of the model for England's literacy and numeracy strategies

6.2. The learning process model for the four initiatives in the United States

6.2.1. Learning contexts

6.2.2. Vertical and horizontal learning

6.2.3. Barrier management

6.2.4. Summary of the model for the four initiatives in the United States

6.3. The learning process model for the two initiatives in New Zealand

6.3.1. Learning contexts

6.3.2. Vertical and horizontal learning

6.3.3. Connections between initiatives in the vertical and horizontal learning dimensions

6.3.4. Barrier management

6.3.5. Summary of the model for the two initiatives in New Zealand

6.4. Learning connections

6.4.1. Learning connections that managed knowledge for practitioners
6.4.2. Learning connections that developed knowledge with practitioners........163
6.4.3. Summary of vertical and horizontal learning connections..................165
6.5. Summary and conclusions.................................................................165

Chapter 7: Policy-Research-Practice Collaborations in New Zealand ..........169

7.1. A vertical policy-research-practice collaboration.................................170
7.1.1. Pre-intervention development needs..............................................170
7.1.2. The fourth-space community of practice........................................172
7.1.3. The third-space community of practice........................................175
7.1.4. Summary of the vertical policy-research-practice collaboration........177
7.2. A horizontal policy-research-practice collaboration............................178
7.2.1. Pre-intervention development needs..............................................178
7.2.2. The fourth-space community of practice........................................180
7.2.3. The third-space community of practice........................................185
7.2.4. Summary of the horizontal policy-research-practice collaboration......189
7.3. Summary and conclusions......................................................................190

PART IV: CONSISTENCY AND CONNECTIVITY TO IMPROVE
INSTRUCTIONAL PRACTICE.........................................................................193

Purpose and approach..................................................................................193

Chapter 8: Discussion and Conclusions.....................................................197

8.1. Condition of the evidence......................................................................197
8.1.1. Patterns of investment into research and development.....................197
8.1.2. Ease of producing outcomes-focused research.................................201
8.1.3. Summary of the condition of the evidence........................................203
8.2. Development and implementation characteristics................................204
8.2.1. Forms of effective school improvement............................................204
8.2.2. Focus on instructional improvement.................................................206
8.2.3. Standardisation...............................................................208
8.2.4. Learning connections....................................................210
8.2.5. Summary of characteristics............................................212
8.3. Accelerating the learning of effective school improvement........212
8.4. Implications.........................................................................215
8.4.1. Conceptual implications for the field of schooling improvement........215
8.4.2. Practical considerations for schooling improvement in New Zealand......217
8.5. Limitations of the thesis......................................................219
8.6. Further research.....................................................................220
8.7. Summary and conclusions...................................................221

Appendices.................................................................................223

Appendix A: Letter for Policy Developers to Comment on Draft Information......223
Appendix B: Descriptions of Five International School Improvement Initiatives....224

References.................................................................................233
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Inquiry process for the theory development</td>
<td>17</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Connections between macro policy contexts and schooling improvement initiatives</td>
<td>26</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Stein and Coburn’s (2005, p.8) framework for analysing research-practice collaborations</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4</td>
<td>A framework for analysing policy-research-practice collaborations adapted from Stein and Coburn’s (2005) framework</td>
<td>116</td>
</tr>
<tr>
<td>Figure 5</td>
<td>A comparative framework to construct learning process models</td>
<td>130</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Learning process model for England’s literacy and numeracy strategies</td>
<td>136</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Learning process model for the four schooling improvement initiatives in the United States</td>
<td>141</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Learning Process Model for the Numeracy Development Project and the SEMO project in New Zealand</td>
<td>146</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Learning connections in three initiatives using a machine bureaucracy</td>
<td>160</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Learning connections in the three initiatives using a professional bureaucracy</td>
<td>162</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Horizontal learning connections</td>
<td>164</td>
</tr>
<tr>
<td>Figure 12</td>
<td>The policy-research-practice collaboration in the Numeracy Development Project</td>
<td>171</td>
</tr>
<tr>
<td>Figure 13</td>
<td>The policy-research-practice collaboration in the SEMO Project</td>
<td>179</td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Policy developers who provided feedback</td>
<td>21</td>
</tr>
<tr>
<td>Table 2</td>
<td>Three categories for selecting effective schooling</td>
<td>54</td>
</tr>
<tr>
<td>Table 3</td>
<td>Evaluative table to select initiatives with the strongest evidence of effectiveness</td>
<td>57</td>
</tr>
<tr>
<td>Table 4</td>
<td>International initiatives with the strongest evidence of effectiveness</td>
<td>59</td>
</tr>
<tr>
<td>Table 5</td>
<td>Constraints analysis for the strong international evidence</td>
<td>67</td>
</tr>
<tr>
<td>Table 6</td>
<td>New Zealand initiatives with the strongest evidence of effectiveness</td>
<td>77</td>
</tr>
<tr>
<td>Table 7</td>
<td>Constraints analysis for the underdeveloped condition of evidence attached to the two New Zealand initiatives</td>
<td>90</td>
</tr>
<tr>
<td>Table 8</td>
<td>Scenario: developing an achievement management system in Mangere</td>
<td>99</td>
</tr>
<tr>
<td>Table 9</td>
<td>Template table for taking descriptive and evaluative notes of the seven initiatives selected for analysis</td>
<td>107</td>
</tr>
<tr>
<td>Table 10</td>
<td>Pen portraits of the seven initiatives selected for analysis</td>
<td>109</td>
</tr>
</tbody>
</table>
I wrote this thesis after doing school improvement work in New Zealand for several years and realising that our efforts alongside other developed countries’ attempts were just not getting the job done quickly enough. The job that I am referring to is helping students from disadvantaged communities leave school with equivalent qualifications and life chances to their counterparts in the leafy suburbs. My involvement in that particular work started almost a decade ago when I left an enormously satisfying position as a school principal of a large inner-city primary school in Auckland, New Zealand’s largest city, to co-ordinate, on behalf of the Ministry of Education (the Ministry) a school improvement initiative in two nearby disadvantaged districts called Mangere and Otara. My role as a change manager for the project was not a solo venture by any means. The project called Strengthening Education in Mangere and Otara (SEMO), was announced in May 1997 by the Minister of Education, Hon. Wyatt Creech, as a three-way partnership between the Ministry, the schools and the communities (Ministry of Education, 1997a, p.246). That announcement brought together local, regional and national Ministry officials with professional and community leaders of the schools as well as researchers and developers interested in helping out.

No one could have prepared us for what we experienced. Not only were relationships frayed but all the groups participating in the project had capacity issues to address. At the outset, the government agency set up to audit schools’ performance, known as the Education Review Office (ERO), indicated that the schools were the problem (Education Review Office, 1996). They certainly had a multitude of problems many of which were distracting them from the core business of teaching and learning (Annan, 1999). Of particular concern was the way they were going about governing and managing their affairs. In the process of dealing with those school performance problems, the other groups in the project started acknowledging that they were also struggling with what and how to improve things.
(Robinson, 2000). Doing school improvement is such hard work (Elmore, 2005; Fullan, 2005a; Rowan, Barnes, & Camburn, 2004). Do we work through relationship issues first or get on with the task and let the relationships sort themselves out? Do we push hard or create incentives, or is it a combination of both? What order do we do things in? How much support do people need to get things done? We found that planning questions such as these were endless and it took time to work through them let alone getting on with improving things.

Coming to the realisation that the project was not going to achieve its end goal of raising the disadvantaged students’ achievement in a hurry caused considerable personal reflection about what to do to help speed things up. Did I continue to toil away in the front-line action of the project? Or did I step back and critically analyse what we were doing compared to the actions of others in similar circumstances to help change the way we do front-line as well as rear-guard school improvement work? I decided on the latter option and also decided that the critical analysis should be doctoral research as opposed to some sort of self or work-based review. There were two reasons for this decision.

The first reason was because I liked to get on with a job when one was given to me. Yet I came to realise that that practical and non-theoretical tendency was not going to help solve the complex problem of underachievement among New Zealand’s disadvantaged students. Indeed, my busy work was adding to the woes and I suspected many others were in the same situation. That practical tendency fits with the New Zealand cultural norm typically referred to as the “No.8 Wire” approach. No. 8 wire is a common type of fencing wire used in New Zealand to fix things around the house or farm. A less than desirable outcome is beside the point. At least the problem is dealt with and further tinkering can be done if necessary. Some commentators on New Zealand culture argue that the approach is an important part of our tradition and it contributes to the success of our nation (J. Hopkins & Riley, 1998). Although this approach may contribute to innovation in some fields, I am not alone in believing that more discipline is needed in areas where complex problems need sophisticated solutions to keep New Zealand vibrant. The current Minister of Finance, Dr Cullen, argued along these lines recently in a speech to the Mathematics in Industry Study Group. The Herald on Sunday reported:
PART I: Setting the Scene

The quintessential-Kiwi [New Zealand] pastime of inventing things with No.8 Wire and a piece of four-by-two is no more than a “little romance” that needs to be put to bed, says Finance Minister Michael Cullen. Dr Cullen told the [group] that to create a vibrant knowledge economy, New Zealand needed a “more disciplined approach… Inevitably this means change. One thing that has to go is the cherished myth of the amateur, the individual who retreats to the garden shed, constructs an unlikely piece of sophisticated equipment and produces something world-beating. (Devereux, 2005, p.1).

Stepping back to complete a critical analysis of what we were doing was a way of moving on from a ‘garden shed’ mentality and bringing a measure of discipline to my thinking. I believed that if my thinking were more disciplined it may influence others in the field to think in a more considered way, which might help speed things up.

The second reason for deciding to conduct a critical analysis of our endeavours, which is also the main reason for deciding to pursue it as doctoral research, was because I needed to improve my writing as a lever for influencing others. Doing practical things and talking ideas into existence came with ease but writing in a theoretical genre was a struggle. Mastering the art of theoretical writing was essential for examining our efforts beyond surface level features in order to reveal the underlying reasons for the slow pace. My quest to gain mastery in theoretical writing does not necessarily mean that everyone in the field has to aspire to that same goal. Some roles may not require this specific skill but I believed that I, in my role as a change manager, did. The role I had undertaken existed largely in between the three world views of educational policy developers, researchers and practitioners. Conducting doctoral research to inform those world views seemed like a good idea at the time. My decision to pursue the research is affirmed by current arguments for practitioners to conduct rather than just consume research (Robinson & Lai, 2006) and for educational leaders to become theoreticians of the practice that they’re expected to lead (Fullan, 2005a). My alignment with these views does not mean that I now believe practice is subservient to theory. To the contrary, the theoretical end-point of this thesis concentrates on the transformation of theory into practice. It indicates that application remains vital to success. It also reminds us to be careful to critically analyse at a theoretical level what we do before and after we apply what we think is the best course of action.
PART I: Setting the Scene
CHAPTER 1
INTRODUCTION

The problem of underachievement among students in schools in economically disadvantaged communities is widespread. New Zealand's problem in this regard is considerable and it is likely to get worse unless things change in the near future. Becoming reading literate, for instance, is critical for disadvantaged students to succeed in a knowledge-rich work environment. Yet recent PISA results for reading literacy of 15-year-olds in New Zealand indicate that there are a large number of students unable to read to the level required to become successful (OECD, 2001). Students from disadvantaged communities are overrepresented in that group. Many of those are Maori (New Zealand's indigenous people) and Pacific in origin (Alton-Lee, 2005). Furthermore, demographic projections indicate that Maori and Pacific students will replace Pakeha (New Zealand European) as the majority student population in schooling within the next three decades (Department of Statistics, 1995). Those achievement and demographic trends have the potential to erode New Zealand's high average in reading achievement in comparison to other OECD countries.

New Zealand government officials as well as educators in schools and in the research and development community are tackling this overarching problem in many different ways. One way is through schooling improvement initiatives and this is the primary focus of this thesis. A schooling improvement initiative is defined as planned interventions designed to raise overall academic achievement of the students in the target classrooms, schools, school districts, regions or country. This definition treats raising academic achievement as the primary purpose of a schooling improvement initiative. The definition of schooling improvement used by Gray, Hopkins, Reynolds, Wilcox, Farrell and Jesson (1999) captures this single focus on raising academic achievement of all students: "If effectiveness describes above- expectation pupil academic performance, improvement is a sustained upward trend in effectiveness. An improving school is thus one which increases its effectiveness over time – the value-added it generates for pupils rises for successive cohorts" (p.5).

Other definitions of schooling improvement initiatives link effectiveness to other outcomes, such as strengthening the school's ability to manage change or to be innovative. A definition put forward by Potter and Reynolds (2002), for instance, defines schooling improvement as: "strategy for educational change that enhances
student outcomes as well as strengthening the school’s capacity for managing change” (p.32). There is an important distinction between that definition and the one being put forward in this thesis. Here, variables such as managing change are not seen as outcomes of schooling improvement. Rather, they are mechanisms attempting to add value to the desired outcome of raising student achievement.

There are numerous lists of mechanisms in the literature which different researchers believe will achieve the desired outcome. Many of the lists are credible in that they synthesise research on how to improve those aspects of schooling believed to contribute to improve student achievement (Harris, 2000; Stoll & Myers, 1998; Stringfield, 2000). Yet such lists are problematic because most of them assume rather than demonstrate the links between interventions and student achievement. That link is often assumed. Whichever mechanisms are chosen for intervention, there is a need to check that they do actually improve student achievement. Causal links between the mechanisms and achievement are not assumed but have to be established in a schooling improvement initiative. They are deliberately set up, analysed and used to inform future developments. Therefore, evidence of student achievement in specific contexts is an essential element of a schooling improvement initiative.

A final important point in defining a schooling improvement initiative is that all the students in the chosen context must be targeted for improved academic performance. If the context is a group of classrooms, then all the students in those classrooms must be included in the initiative. The same applies for an initiative involving a group of schools in a region or across an entire country. There is nothing indirect about a school improvement initiative. It directly adjusts school and teacher practices so that all students can benefit. This means some well-known pull-out interventions, such as Reading Recovery, were excluded from consideration of successful schooling improvement initiatives.

1.1. The Research Problem

The research problem that this thesis is intended to help solve is the slow pace of learning effective school improvement. The problem has many aspects to it. I believe that it is best to try and solve those aspects of the problem that are within the control of those people involved in school improvement. Going beyond what they can control is not likely to generate any enthusiasm or change. For instance, schooling
improvement initiatives have been part of broader reform to improve schooling beyond equity issues, such as improving governance and health and safety of schools’ staff, students and buildings. In discussions about school reform trends between United States and French researchers, Susan Fuhrman (2002) noted that there has been relentless endeavour in this regard through waves of reforms over 15 years. She stated: “They come steadily and rapidly, wash over us, and sometimes leave only traces over the long run... They are united by a number of themes, but they typically had different primary champions, took different points of departure and were addressed at different targets” (Fuhrman, 2002, p.1). Fuhrman’s take on overall school reforms conjures up a messy state of affairs, a point Cohen (2002) put succinctly at a public policy forum in Washington, DC: “Education is an incredibly complex institution and undertaking. We live in a messy world” (Cohen, M., in Federal News Service, 2002). The broader messy picture that these scholars conjure up is not something that participants of schooling improvement initiatives can control and they are highly unlikely to get excited about sorting it out as their next challenge.

An aspect of the slow pace of improvement which is relevant to and more likely to enthuse participants working in the field is that learning how to do effective schooling improvement is extremely difficult. Fullan (2005a) highlights this part of the problem in his argument that sustained improvement is reliant on leaders who routinely theorise about practice and connect with relevant parts of the system: “the agenda is exceedingly complex and demanding... you can’t be a system thinker in action if you don’t know what the action part looks like and feels like. Learning by doing has never been so thoughtful and so challenging” (p.x). Part of this challenge in relation to schooling improvement groups is that the ‘doing’ groups, such as policy developers, researchers and practitioners, are not used to learning from one another. School reform researchers indicate that traditional connections between those three groups have not tended to be learning encounters (D. Hargreaves, 2003a; Hill, 2001; Spillane, Reiser, & Reimer, 2002). Rather, encounters have typically been a hand-over of “a set of axiomatic truths” from policy developers and researchers to practitioners (Hill, 2001, p.312).

Two major assumptions underpin hand-over encounters from policy developers and researchers to practitioners. The first assumption is that the practitioners understand what is being handed over to them. Leading researchers investigating sense making within school reforms have found that assumption to be flawed
Chapter One: Introduction

(Coburn, 2003; Spillane et al., 2002). Misunderstandings commonly occur at the point of transfer because insufficient investment is made in ensuring that the receiver has fully understood what is being handed over. The second assumption is that after a hand-over, practitioners have the skills and knowledge to work out independently how to implement useful ideas. Once again, that assumption appears to be flawed.

Hargreaves (2003a) acknowledges that practitioners have mainly been left to themselves to figure out what to do and what not to do as a follow up to what they have received. He describes how teachers have gone about fine-tuning what they are given: “Teachers, like artisans, work mainly alone, cobbling together ideas and materials out of which, through repeated tinkering, they devise strategies and routines to making teaching and learning work effectively in variable conditions” (p.34). However, Hargreaves (2003a) as well as other school improvement researchers are questioning this form of independent knowledge “cobbling” or craft knowledge in terms of reliable knowledge management (Alton-Lee, 2003; Datnow & Stringfield, 2000; Stein & Coburn, 2005). Datnow and Stringfield (2000) suggest that if it were the lives of passengers on aeroplanes and not the academic lives of disadvantaged students we were talking about, there is no way practitioners would be left to their own devices.

The two assumptions about practitioners understanding things and being highly skilled have their origin in part in a schooling culture that has typically placed high value on politeness and friendship. Politeness among practitioners at the expense of asking critical and challenging questions to check for understanding is a recognised feature of the schooling culture internationally (Ball & Cohen, 1999; De Lima, 2001) and in New Zealand (Annan, Lai, & Robinson, 2003; Robinson & Lai, 2006). One source of the politeness appears to be traditional values within the educational community about what constitutes a good and respected professional educator. An important value in that regard is that professionals know things so their conversations start from a position of knowing rather than inquiring into what is not known. So when teachers engage in a professional conversation about teaching reading, for instance, they are more likely to talk about what they know than to reveal what they don’t know.

Furthermore, they are unlikely to challenge what their colleagues say they know because of the influence of professional friendship ties. De Lima (2001) found that many teacher-networks were formed through strong personal ties or friendships
rather than professional ties. She also found that friendship-focused networks of teachers detract from learning: “Very strong ties may discourage the acquisition of new information and social identities and thus retard innovation and change” (p.208). De Lima advocates for greater challenge and respectful resolution of conflicting ideas among practitioners to mediate between professional and personal ties. Yet those two elements of mediation can cause professional discussion to become personal and hence to be avoided.

In summary then, the research problem is the slow pace of learning effective reform practices. The particular aspect of that problem of interest is the way policy developers and researchers have tended to engage in hand-over encounters with practitioners within a polite and friendly schooling culture. So there are two aspects to the research problem. One is a direct focus on hand-over encounters between policy developers, researchers and practitioners. The second is the consequential impact of those encounters on the culture of schooling. There are many ways of thinking about investigating this two-tiered problem. One way is to deal with both tiers at the same time but that was considered unmanageable mainly because the schooling culture is such a broad concept. Another way is to investigate the encounters between the three groups and to use the findings as a lever to suggest ways of influencing changes to broader culture of schooling. That approach seemed more manageable and it also concentrates on accelerating elements that are within the control of those leading schooling improvement initiatives and that have already been shown to be effective.

1.2. Purpose of the Thesis

The purpose of the thesis is to examine the effectiveness of schooling improvement initiatives in New Zealand in order to help accelerate the learning of effective reform practices. It is a complex purpose in that it presumes a concept of effectiveness and demands an in-depth understanding of how effective initiatives have been developed and implemented.
1.2.1. Effectiveness

Considerable work is done at the beginning of Part II to develop a defensible approach to evaluating effectiveness. An effective schooling improvement initiative will be defined in this thesis as one that makes a positive and statistically significant impact on the overall academic achievement of the targeted students. Comparative achievement data generated from scientific research is recognised as useful information for reform participants to find out whether their efforts have been successful (Borman, Hewes, Overman, & Brown, 2003). Two common ways of generating such data are calculating percentage gains and calculating effect size gains. Those types of statistical calculations can show positive improvements in student achievement as distinct from a negative or nil improvement (Coolican, 1999). Positive effect-sizes and other measures of statistical significance can represent real improvement in student learning or fluctuations that might occur by chance inside a classroom. When the probabilities of getting a result by chance are low, then it is statistically significant at a given level\(^1\). By calculating the statistical significance of student learning, schooling improvement participants build up evidence that gives them reason to either go back to the drawing board, because the improvement was too small, or get excited about the impact of their initiative.

An effect-size, for instance, calculates the size of the effect on student learning of an intervention using a scale with 0 representing nil impact. A score below 0 constitutes a negative effect and a score above 0 shows a positive effect. Hattie (1999) used effect-size measures in a meta analysis of 180,000 studies about schooling innovations and found the most positive schooling innovation to be teacher feedback to students (\(d = 1.13\)). Retaining students for an extra year at the same year level was found to be the most negative innovation (\(d = -.15\)). By using a common effect size indicator, such as Cohen’s d, equivalent effect sizes can be calculated (Lipsey & Wilson, 2001). Equivalent effect size measures are particularly useful for comparing the success of different initiatives.

The definition of an effective schooling improvement initiative that is put forward in this thesis is contentious. Two challenges amongst others are inevitable. The first

---

\(^1\) Two standard cut-off points help classify the statistical significance of a result. One is when the probability of chance or error is reduced to less than 0.05, also known as the five percent significance level; the other is when the probability is reduced to less than 0.01.
likely challenge is to the underlying assumption that schooling can make a difference to student learning. That assumption supports an overarching theory in much of the schooling improvement literature in New Zealand that schooling and, in particular, teaching can make a positive difference to student learning (Alton-Lee, 2002, 2003, 2004; Hattie, 1999; Lai, 2005; Lai, McNaughton, MacDonald, & Farry, 2004; Lai, McNaughton, MacDonald, Amituanai-Toloa, & Farry, In Press; McNaughton, Lai, MacDonald, & Farry, 2004; Phillips, McNaughton, & MacDonald, 2001; Robinson & Lai, 2006; Timperley, Phillips, & Wiseman, 2003). Underpinning their arguments is an implicit theory that teaching, supported by the wider schooling system, can make a much better impact on the underachievement problem than it is currently doing. They are not advocating massive interventions or legislative change. Rather, they are simply getting on with designing, implementing and evaluating evidence-informed interventions. Collectively, they are testing the assumption that causal links between social conditions and performance are inevitable. The theory is that the link between student achievement and social background was socially constructed and, as inconceivable as it may seem to some educators, the link can be socially dismantled. So energy is best spent working out ways to dismantle the link rather than defending its existence.

However, some international and New Zealand researchers remind those in support of the implicit theory for improvement described in the previous paragraph that schools alone, and in particular teachers, cannot dismantle long standing inequalities such as the link between low achievement and disadvantage (Harker, 2004; Hawke et al., 1996; Nash, 2003; Nash & Prochnow, 2004; Thrupp, 2001; Tunmer, Chapman, & Prochnow, 2004). Their arguments suggest the problem is much wider than the teacher or the school. Broader social issues have to be addressed for schooling to significantly impact on the academic achievement of students from disadvantaged communities. Issues include better health and welfare services and adequate community and family resources to support student learning. Harker (2004) suggests that things are compounded in large urban districts. He claims, for instance, that high status schools attract high ability students from surrounding districts and that has a detrimental affect on low status schools’ achievement profiles. The complex matrix of constraints makes it incredibly difficult for schools to address the inequalities that exist. Nash and Prochnow (2004) aptly sum up this counter theory that a much broader improvement agenda is necessary to comprehensively address the entrenched
inequalities: “In the face of all the evidence, it is unrealistic to expect that the attainments of middle-class and working-class families can be equalized, as some speakers within this broad discourse assert, as a result of pedagogic action by the school” (p.189).

Both arguments about making progress and the difficulty of doing so are true. Schools can make a significant difference in disadvantaged communities and improved social systems, such as health and welfare, would provide a better social foundation for academic learning. However, this thesis does not attempt to find solutions for systems beyond the schooling system. Rather, it focuses on solutions that are within the control of policy developers, researchers and practitioners participating in or influencing schooling improvement initiatives. The aim, therefore, is to identify those conditions that can be manipulated by those three groups through schooling improvement initiatives and look to inform the thinking around how those conditions can be improved. That aim does not fully respond to Nash and Prochnow’s point in the previous paragraph about schooling solutions not being able to address the inequalities. There is no intention to do so because the central issue is about raising achievement and not equality. That would require research into whole-of-government approaches. Such a task is well beyond the scope of this research and, I anticipate, of little value to the specific concerns of schooling improvement initiatives.

The second challenge to the definition is that quantitative measures, such as overall effect-sizes and percentage gain calculations are unhelpful to schooling improvement participants, particularly teachers thinking about changing their instructional practices. The concern is that aggregated quantitative data tells teachers very little about whether the intervention will work in their classroom. That point is not in dispute. School teaching has long been recognised as a highly contextualised task (Kliebard, 2002; Stringfield, 2000). However, aggregated achievement information and context-specific decision making are not necessarily incompatible. Compatibility can be achieved by using aggregated achievement information for two purposes; management and formative assessment. In terms of management, it is important for reform sponsors, school leaders as well as teachers to gain an overall indication of the probability of success of the interventions in which they are investing. It is better for teachers and school leaders to use this evidence as a guide to select or eliminate an intervention rather than to take a data free approach. Without
this guidance, practitioners are reliant on the general attributes of interventions, such as levels of support and implementation costs, to guide their decisions.

In terms of formative assessment, aggregated achievement information can be used by teachers to check for causal links between what they are doing and how well their students are achieving. For that to happen, a critical aspect of aggregating achievement information beyond a classroom or school is to ensure that teachers and those helping them can disaggregate the information in order to track individual students’ progress. Harker (2003) reinforces this point by recommending that researchers studying disadvantaged students should collect data about the students’ backgrounds and academic achievements at the individual level in order to make valid claims of improvement:

A lot of information is lost when data is aggregated at the school level, and when gains are small in aggregate it is clear that some pupils will have made gains, many will have stayed much the same, and some will have gone backwards. We really need to know about the trajectories of individual pupils and more about their individual backgrounds in order to try to explain “what works with whom, and why?” (p.246)

Numbers on their own do not explain the meaning of the practices and processes that produced them. Numerical calculations tell us what works and what does not work. Without appropriate controls, they cannot explain how or why things work or do not work (Nash, 2002). Qualitative narratives, such as ethnographies and case studies can. Those types of studies uncover the vital ingredients of the initiatives. They provide a depth of meaning about why an initiative is or is not working. That is not to say that qualitative narratives are preferred over quantitative effect size measures. One without the other misses the full picture. Both types of measure make it possible to construct a programme theory that explains what works for whom in what circumstances (Pawson, 2002). The point here is that numbers are particularly useful for selecting successful initiatives and narratives are useful to find out why they succeeded.
1.2.2. Development and implementation of schooling improvement initiatives

An in-depth understanding of how schooling improvement initiatives have been developed and implemented requires an analysis of existing initiatives. In order to find out which initiatives in New Zealand were most effective and understand what was happening in those initiatives, a decision was made to broaden the investigation to include an analysis of effective initiatives from other countries with similar systems level characteristics. That decision enabled useful comparisons and a much deeper understanding than would have been possible by just analysing initiatives operating in New Zealand. A critical characteristic was that the comparison countries were developed countries and had:

a) market economies;

b) democratic governance, and

c) a government investing in school improvement initiatives for students living with disadvantage.

Countries with market economies that are democratically governed define a ‘developed’ country as this term is used by the Organisation for Economic Co-operation and Development (OECD) (OECD, 2003b). Membership and non-membership of the OECD generally distinguishes between developed countries and developing countries. The OECD has 30 member countries, 26 of which have established market economies and democracies. Many of those developed countries have placed the task of improving schooling for students living with disadvantage high on the social policy agenda over the last 15 years (Fuhrman, 2002). It is those countries that I believe are best suited to generate comparisons relevant to improving New Zealand’s school improvement efforts. Literature about schooling improvement in developing countries was therefore set aside.

---

2 There are exceptions, for instance Brazil and India fit the definition but are not members of the OECD and Turkey and Portugal are members of the OECD but their overall performance suggests they are developing countries.

3 Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Slovak Republic, Spain, Sweden, Switzerland, Portugal, Turkey, England, United States.
1.3. Research Questions

Three research questions listed below help focus this thesis onto evidence of effectiveness and what is happening within effective schooling improvement initiatives.

(i) What are the most effective national and international examples of school improvement?

(ii) What is the condition of the evidence base for making claims of effective schooling improvement?

(iii) What can be learned about developing and implementing effective school improvement from international and national initiatives with the strongest evidence of effectiveness to inform New Zealand’s school improvement movement?

Those three questions set the boundaries for this thesis. The boundaries are intentionally tight in order to achieve considerable depth in understanding about what was happening within the seven initiatives with the strongest evidence of effectiveness selected for analysis. The tightness led to an inductive inquiry process which involved much analytical iteration to describe development and implementation characteristics and explore interrelationships between the characteristics within the seven initiatives.

One important finding among several that came out of the iterative inquiry process was the importance of learning connections between policy developers, researchers and practitioners participating in the seven initiatives selected for analysis. Learning connections are about the collective pursuit of knowledge among participants of schooling improvement initiatives to transform useful theoretical and practical ideas into effective practice. They are also about the transfer of that knowledge into teaching practices that identify and solve achievement problems as quickly and as early as possible. Connections in themselves are not a form of learning. Rather, they are one condition that can be useful for learning. Connecting to socialise or to find
out about schooling issues other than solving achievement problems is not the sort of
connections that this theory development is about.

It is those connections that are central to the theory for improvement developed in
this thesis. By ‘theory of improvement’ I mean a set of linked ideas to develop and
implement an intervention to raise achievement among a targeted group of students.
The main reason for selecting learning connections ahead of other elements of school
reform in the theory development is because they are critical to accelerate the
acquisition and use of knowledge to identify and solve achievement problems. That
knowledge management component is central to my argument for accelerating the
learning process. An assumption underpinning my argument is that existing
knowledge about school improvement coupled with the knowledge generated from
this thesis is sufficient to get much closer to solving the underachievement problem
than past efforts have managed to achieve. Another assumption is that there are
already communities of reformers putting the theory into practice.

The research questions also indicate what the thesis is not about. Broadening the
investigation to find out how other groups like parents and students connect with
policy developers, researchers and practitioners was given serious consideration but
was found to be beyond the constraints of this particular study. That is not to say that
those other relationships are not important. They are critically important. For
instance, a strong educational connection between schools and families make an
important contribution to solving achievement problems among disadvantage students
(Bishop, Berryman, Tiakiwai, & Richardson, 2003; Ministry of Education, 2006e).
An educational connection implies that the school knowledge about teaching and
learning will inform ways that families support their children’s learning at school. It
also implies that family knowledge of their children’s lives outside school will inform
teaching practices in school. Based on this conceptualisation of schooling, the
expectation of a “significant difference” in the overarching theory is contingent on
learning connections between schools and families. Other researches are
investigating those connections and they are producing a growing pool of evidence
supporting pedagogies connected to the lives and cultures of the students being taught
(Bishop et al., 2003; Macfarlane, 2004; Ministry of Education, 2006e; L. Smith,
1999). It is critically important that those sorts of connections with non-professional
players are considered alongside the sorts of professional connections explored in this
thesis.
1.4. Methodology

Figure 1 summarises the inductive inquiry process used to answer the three research questions. Overall, it is a theory development process which creates strong links between evidence of effectiveness and development and implementation characteristics of effective schooling improvement initiatives. An illustration of the process is important as it is so central to the methodology. At the outset, the researcher is more interested in finding evidence of effectiveness attached to initiatives than learning about the characteristics of them. That avoids sifting through considerable amounts of information that would be set aside in the selection process. Once the most successful initiatives are identified and selected, the researcher can concentrate on critically analysing information about them to discover the common characteristics relevant to accelerating the learning of effective reform practices. The ordering of the inquiry process to that point assures causal links between evidence of effectiveness and development and implementation characteristics. The researcher is thereby assured that the theoretical discussion at the end of the study has a strong evidential basis.

- Review of evidence of effectiveness on schooling improvement
- Selection of schooling improvement initiatives with strongest evidence of effectiveness
- Critical analysis of development and implementation characteristics of the selected initiatives
- Theory for improvement based on characteristics of initiatives with the strongest evidence of effectiveness

Figure 1. Inquiry Process for the Theory Development.
1.4.1. Ethics

The methods used to implement the inductive inquiry were determined to a large extent by the ethical constraints attached to my official status as a Ministry employee. The nature of my position constrained the type of research that I was able to do. There were three ethical constraints in particular that determined the methodology that could be used. The first ethical constraint related to issues of power and control. Had I made direct contact with researcher-developer or practitioner-participants of the initiatives in the data gathering stage, their responses may have been tempered in an effort to safeguard relationships or future contract opportunities with the Ministry. The second ethical constraint related to the need to make sure that information that I had access to but was not intended for the public arena was safeguarded. There was information held within the Ministry’s school improvement databases that could be detrimental to individuals and/or organisations if it were publicised, such as human resource management information. Some of that information also resided in my head and it was particularly important for the individuals and/or organisations concerned that I did not divulge it in this research project.

The third ethical constraint was to remain loyal to the civil service code of conduct as a Ministry employee (Ministry of Education, 2006b). A particularly relevant statement in the code in regards to this research is public servants’ involvement in public debate: “Public servants should ensure that their contribution to any public debate or discussion on such matters [Government policy] is appropriate to the position they hold, and is compatible with the need to maintain a politically neutral Public Service” (Ministry of Education, 2006b, p.6). My position as a change manager in the Ministry’s Schools Monitoring and Support division was interesting in regard to this statement. It was expected to influence change among all the groups involved in schooling improvement initiatives, which included policy developers as well as researchers, developers and school leaders (Ministry of Education, 1997a). Hence, critical analyses of those three groups in the public debate about schooling improvement in New Zealand were appropriate to the position. In terms of political neutrality, there was no risk of criticising government policy through my contribution to public debate because there is no formal schooling improvement policy to speak of in the area that this thesis explores, i.e. the links between outcomes-focused evidence and development and implementation characteristics attached to school improvement
initiatives. Given that senior Ministry officials approved and funded the research, it was probable that they were anticipating that the information would help formulate such policy.

1.4.2. **Data gathering and accuracy checks**

I employed two key strategies to deal with the ethical constraints attached to being a civil-servant researcher. The first strategy was to restrict data gathering methods to the analysis of publicly-available documents. This strategy immediately dealt with the issues of power and control and sensitive information attached to my relationships with researcher-developers and practitioners. The strategy meant that there was no data gathering through those participants. For instance, there were no interviews or observations to find out how researcher-developers and practitioners were connecting with one another in the seven initiatives selected for analysis. Data gathering of that nature involves direct relationships with participants and these were not possible because of my position as a Ministry official. Instead, the inductive inquiry involved a considerable amount of analysis of already published descriptive and evaluative information. Different analytical techniques and tools had to be employed because the nature of the analysis tasks kept changing as the theory was developed. Rather than collating and reporting all the techniques and tools here, I decided to outline the relevant methodological approaches at the beginning of each of the four parts. Consequently, the introduction to each part is divided into two sections. The first section is the ‘purpose’, which links the contents of the part to the research questions and the second section is the ‘approach’ which outlines the methodologies used to complete each part.

There were some guiding principles that I adhered to in moving away from a more traditional “Methods” chapter. An overarching principle was that the approach outlined at the beginning of each part was consistent with the inquiry process summarised in Figure 1. That made sure that Parts I, II, III and IV comprised a whole study and were not presented as discreet studies on their own. Another important methodological principle was to identify and, where necessary, adapt analytical frameworks that other researchers have already developed and found useful in school improvement research. This principle proved to be invaluable in Parts I, II and III to produce findings that would usefully inform the theory development in Part IV. In
Part I, Spillane et al’s (2002) conceptualisation of school improvement as a nested learning arrangement helped form a basic model for investigating school improvement activity across various contexts. Then in Part II, the approach to developing evaluation criteria to select initiatives with the strongest evidence of effectiveness was greatly assisted by criteria developed by Borman et al (2003). Also in Part II, Robinson and Lai’s (2006) constraints analysis technique was useful for analysing the reasons why evidence varied in quality nationally and internationally.

In Part III, the basic model developed in Part I was used to help develop the comparative framework and to create the learning process models. It was ideal because its origins were based on learning principles for children (Bronfenbrenner, 1979). Furthermore, some school improvement researchers are now suggesting that there is little, if any, difference in learning processes suitable for adults and children (Ministry of Education, In Press). Although the learning process models developed in Part III produced some interesting findings, they only touched on a really interesting phenomenon of policy-research-practice collaborations in New Zealand. I considered those collaborations worthy of further investigation and searched for a suitable analytical tool to investigate them. The search found a conceptual framework developed by Stein and Coburn (2005) to be most suited to the task. They had designed it to analyse research-practice collaborations. However, it was easily adapted to analyse policy-research-practice collaborations in New Zealand. It also helped get some objectivity into my analytical thinking about projects that I had been closely involved in.

Another strategy to deal with the constraints of being a civil servant-researcher was completing accuracy checks of my analyses. This strategy not only mitigated the inclusion of inaccurate and sensitive information but it also checked for personal biases and prejudices. This was particularly important in light of my personal involvement and interest in the success of the SEMO project. An overall accuracy check of the contents in each part was completed by re checking the source documents to see if there were any alternative explanations that had not been considered. Alterations were made when disconfirming evidence was found.

In the case of the information about New Zealand’s macro policy context and school improvement initiatives, additional checks of accuracy were achieved by seeking feedback from a group of policy developers. I considered this strategy appropriate because my relationship with policy developers was not heavily laden
with power and control issues, as it was with practitioner and researcher-developer participants. I was part of the broader Ministry team involved in policy development. The process used to select the group was to identify local, regional and national officials most closely associated with the two projects and ask them to provide feedback. Twelve officials were identified (see Table 1). They were sent an electronic and paper copy letter with attached draft information on which to comment (See Appendix A). Nine of the 12 officials responded within the one month timeframe given to comment. Several of those officials chose to continue commenting on subsequent drafts through to the final draft submitted for examination (National senior advisor for numeracy, national Manager with oversight for cottage industry initiatives, national senior policy analyst, local schooling improvement co-ordinator, manager of the Best Evidence Synthesis project). Table 1 outlines the role of each official that responded, the project that they were associated with and their relationship with the project. The nine respondents represented three officials checking the accuracy of the information in relation to the Numeracy Development Project, five checking it in relation to the SEMO project and one checking the accuracy of both projects.

Table 1. Policy Developers Who Provided Feedback.

<table>
<thead>
<tr>
<th>Policy Developers Who Provided Feedback</th>
<th>Project association</th>
<th>Relationship with the project</th>
</tr>
</thead>
<tbody>
<tr>
<td>National manager with oversight for all national professional development programmes</td>
<td>The Numeracy Development Project</td>
<td>Senior official overseeing funding approval for the Numeracy Development Project and active involvement in monitoring the project</td>
</tr>
<tr>
<td>National senior advisor for numeracy</td>
<td>The Numeracy Development Project</td>
<td>National co-ordination of funding, monitoring and support to the Numeracy Development Project</td>
</tr>
<tr>
<td>Local schooling improvement manager</td>
<td>The Numeracy Development Project</td>
<td>Local co-ordination of monitoring and support to the Numeracy Development Project in the district of Manurewa</td>
</tr>
<tr>
<td>National manager with oversight for all cottage industry initiatives</td>
<td>The SEMO project</td>
<td>Senior official overseeing funding approval for the SEMO project and active involvement in monitoring the project</td>
</tr>
<tr>
<td>National senior policy analyst</td>
<td>The SEMO project</td>
<td>Active involvement in monitoring the SEMO project</td>
</tr>
</tbody>
</table>
National manager of the monitoring and support division | The SEMO project | Senior official involved in funding approval, monitoring the project and managing the local officials involved in the project
---|---|---
Regional manager co-ordinating support for at risk schools | The SEMO project | Co-ordination of monitoring and support for schools at risk in the districts of Mangere and Otara.
Local schooling improvement co-ordinator | The SEMO project | Local co-ordination of monitoring and support to the SEMO project
Manager of the Best Evidence Synthesis Project. The project is synthesesing useful evidence to inform teaching and learning in New Zealand early childhood settings and schools | Both | The syntheses draw on various interventions including the Numeracy Development Project and the SEMO project

The policy developers’ feedback was received in written form via email or handwritten comments on paper drafts. The feedback was collated and then each point was cross checked with the descriptive information in the notes and the draft text in this chapter. When they made a point that was relevant but not covered, I integrated it into the text and acknowledged them for it. That happened when the policy developers shared their extensive knowledge of the macro policy context. When they disagreed with a particular point, there were several eventualities. One was to agree with them, make the change in the text and insert the source of the evidence backing up the alteration. Sometimes that source was the policy developers’ explanation and other times it was an official document or evaluative report. Another eventuality was to debate the disputed point and make alterations iteratively until we agreed that the description was more accurate. That happened when I either did not understand what they were saying or when I believed there were inconsistencies in their arguments. Either way, evidential references were sought and inserted to support the final text.

A point of finality was reached with regard to the accuracy of the information related to the New Zealand context after the policy developers had been given ample opportunity to discuss my responses to their critical feedback in subsequent alterations to the text. It was a process of checking that I correctly understood their critical feedback by resolving disagreements through learning conversations and representing the information to ensure their views were interpreted correctly. A learning conversation is defined as: “Learning about the meaning and quality of each
other's views" (Robinson & Lai, 2006, p37). So while the process helped me check my understandings, it was also giving the policy developers an opportunity to think about their understandings. An intended consequence of this approach was to get important findings in the thesis into the minds of busy national policy developers early rather than relying on them reading the published document. That endeavour mainly relates to reducing lag time in getting useful ideas on the national policy agenda.

The rigorous process of checking findings and seeking agreement was important not only to meet the requirements in the civil service code of conduct. It also served the purpose of increasing the validity of the findings. In a recent book published by Robinson and Lai (2006) intended to help educational practitioners conduct quality research, the authors make the point that a checking process is a critical methodological principle for increasing the validity of findings. They distinguish between simply gaining agreement and going on to increase validity:

It is important to note that the purpose of such checking is to increase validity and not simply to gain agreement. As Denis Phillips, the philosopher of social science, has noted, it is possible to find people who agree with a claim regardless of its validity (Phillips, 1987). For example, a principal may claim that bullying in the playground is decreasing, and his teachers may agree with him. However, both the principal and the teachers may be wrong. Gaining agreement will only increase validity if it is the result of careful consideration of the grounds on which the claims are made. (Robinson & Lai, 2006, p.63)

That principle of checking with participants for accuracy has its origin in a type of validity check called "respondent validation" put forward by Hammersley and Atkinson (1995). That particular form of checking anticipates that the participants attached to the writing may have additional information unknown to the researcher or may have experienced things differently to what is in the draft text. The principle of respondent validation is introduced here because it was used to check all the information relevant to New Zealand, which is spread throughout the thesis.
1.5. Organisation of the thesis

The thesis is organised in four parts.

Part I sets the scene for the investigation into effective school improvement. Chapter One outlines the overarching underachievement problem facing the field of school improvement and then focuses on the specific research problem that this thesis addresses, that being the slow pace of learning effective reform practices. The purpose of the research is outlined along with the research questions and an overview of the methodology. Central to the methodology is critical analysis of publicly available documentation. As a range of broad and deep analyses of relevant documentation were completed throughout the thesis, explanations of the methodology for part’s I, II, III and IV are included at the beginning of each of those parts. Chapter One is rounded off with a brief description of the seven school improvement initiatives selected for analysis in this thesis.

Chapter Two provides information about the macro policy contexts for England, the United States and New Zealand as those are the three countries in which the seven initiatives operated. By macro policy context I mean the ways that governments in different countries have exerted influence on schooling improvement initiatives in those three countries. That influence is important because it often determines what can and cannot be done in schools that are partly or fully government funded and subject to related regulations. The aspect of government policy that is of particular interest in this thesis is the way that the policy community connects with the research and development and practitioner communities involved in schooling improvement initiatives.

The policy community is made up of politicians, national policy developers and local officials working on behalf of the government to solve problems of the day. The research and development community is made up of researchers, developers,
researcher-developers, consultants, co-ordinators and advisors with the skills and knowledge to help schools intervene into achievement problems. The practice community is made up of “practitioners” involved in the act of teaching or having direct management oversight of teaching within schools; teachers, middle managers (under various titles such as assistant principal, deputy principal, associate principal and director) and principals. Including middle managers and principals alongside teachers in the definition of “practitioner” recognises teachers and students may be alone physically in their classrooms but their practices are not disconnected from those operating in the multiple layers surrounding them (Alton-Lee, 2002; Coburn, 2005). Together those three groups form the broad professional community involved in school improvement and the interactions between them are particularly important in learning effective reform practice.

In order to describe the macro policy contexts in relation to schooling improvement initiatives, I created a model to show the relationship between those two constructs. The model is presented in Figure 2 and it has four contexts; a macro policy, a schooling improvement initiative, a school and a classroom. The reason for selecting those four contexts is because they are places where policy developers, researchers and practitioners typically tend to interact with one another in relation to schooling improvement initiatives. Teachers and students are at the centre of the system because there is general accord at least within New Zealand’s policy community that effective teaching practices are the best systemic lever to lift the overall academic performance of students in disadvantaged communities (Alton-Lee, 2002, 2003, 2004). So the teachers and students are nested within the school and community, which are in turn nested within a schooling improvement initiative. All three of those inner contexts are nested within the politics and social structures of a country, which includes the macro policy context set by the government of the day.

The idea that schooling is a nested arrangement originates from Bronfenbrenner’s (1979) conceptualisation of the ecological environment surrounding human development. He conceived a child’s learning environment as: “a nested arrangement of concentric structures [micro, meso, exo & macrosystems], each contained within the next” (p.22). Schooling as a nested arrangement is now a recognised phenomenon within the school improvement literature. In a study about teachers making sense of standards-based reforms and implementing those reforms in the United States, Spillane, Reiser and Reimer (2002) found that:
In general, implementing agents' [teachers'] work is nested in multiple organisational contexts simultaneously (p.409). Implementation [teaching] practice is not simply a function of an individual agent’s [teacher’s] ability, skill, and cognition; rather, it is constituted in the interaction of administrators, teachers, students, and their situation in the execution of particular tasks. Hence the activity system as distinct from the individual teacher’s or administrator’s knowledge structures becomes the appropriate level of analysis. (p. 412)

Spillane et al’s (2002) idea about schooling improvement being a nested arrangement and the activity system being a useful unit of analysis were critical design elements used to create the model.

![Figure 2. Connections between Macro Policy Contexts and Schooling Improvement Initiatives.](image)

The vertical arrow in Figure 2 indicates that government influence is exerted across all the contexts involved in schooling improvement. One New Zealand researcher believes that government influence within education has historically manifested itself in two forms (Stephenson, 2000). The first form is by government
responding to demands from reform groups to solve problems. The second form is through government’s pro-active management of problems so that demands are minimised. These two forms of influence imply that state intervention is always active. For instance, a government can provide support to particular school reform groups demanding help or it can use its bureaucracy to intervene directly. However, state influence can also be passive. By not intervening, the state can influence other support organisations, such as universities, to provide support in its absence. Whether the influence is active or passive, it represents the macro policy context within which activities such as school improvement operate.

Part II searches out and critically analyses the evidence of effectiveness attached to the seven school improvement initiatives found to have the strongest evidence. In the introduction to Part II criteria are developed for selecting initiatives with the strongest evidence of effectiveness for further analysis. The criteria are then used in Chapter Three to select five international initiatives for analysis. The five initiatives include England’s literacy and Numeracy initiatives and four initiatives in the United States; Direct Instruction, Success For All, The School Development Programme & a district-wide reform in New York District #2. After explaining and defending the selection process, an analysis of and explanation for the overall condition of the international evidence base is provided. Chapter Four repeats the evidential selection and analysis process used in the previous chapter for New Zealand school improvement initiatives. In this case, the criteria helped select the Numeracy Development Project and the SEMO project for further analysis. Interesting comparisons are made about the condition of the evidence bases attached to the initiatives in the three different countries.

Part III is an analysis of development and implementation characteristics of the seven initiatives relevant to learning effective school reform practices. The introductory section to Part III outlines the design of two important analytical frameworks developed to understand what and how learning was encouraged in the initiatives. The first tool is a comparative framework which was designed to create models of the characteristics relevant to the learning processes used in the seven initiatives. Evidence to support the design of the comparative framework and a description and explanation of the framework itself are presented in Chapter Five. The learning process models are then presented and explained in Chapter Six along with an analysis of the learning connections occurring between the three main groups
represented in the learning models, i.e. policy developers, researcher-developers and practitioners.

Findings from chapters five and six represent relatively broad insights into the learning processes that were being used in the seven initiatives. Hence a need to conduct a further investigation to get deeper insights into useful learning processes. By “deep” I defer to Coburn’s (2003) definition of the term: “getting beyond surface structures or procedures (such as changes in materials, classroom organisation, or the addition of specific activities) to alter teachers’ beliefs, norms of social interaction, and pedagogical principles as enacted in the curriculum” (p.4). Then in Chapter Seven, some interesting collaborations among policy developers, researchers and practitioners in the New Zealand initiatives that were not found in the international initiatives are investigated in some detail. That final investigation necessitated the design of another framework, this time a conceptual one. Collectively, the three investigations in Chapters Five, Six and Seven provide both broad and deep perspectives of relevant development and implementation characteristics to accelerate the learning of effective reform practices in New Zealand.

Part IV shifts the focus from findings to discussion and implications and it also completes the final requirements of the thesis including its limitations, further research and conclusions. This part involves only one chapter, Chapter Eight, which begins with a two-part discussion. The first part of the discussion explains why evidence of effectiveness is so much stronger overseas. A critical point made is an evaluative tendency in New Zealand to assist rather than assess school improvement initiatives. The second part of the discussion concentrates on three interdependent characteristics common to all seven initiatives; a focus on instructional improvement, standardisation and learning connections. Replication of those three characteristics into initiatives that are struggling to get traction is offered as one way of speeding up the learning process. I conclude the discussion by arguing that such a strategy is useful but not sufficient to accelerate the learning process as much as is necessary to achieve substantial in-roads into the underachievement problem. A second strategy of connectedness between initiatives to solve problems is offered, which levers off the policy-research-practice collaborations outlined in Chapter Seven. The discussion points culminate into a theory of consistency and connectivity to improve instructional practice. Theoretical implications for the field centre on current thinking about professionalism versus prescription and the sorts of connectors that are
necessary to create a more linked school improvement landscape. Practical implications for reformers in New Zealand suggest that whilst much evaluative work is needed to ensure that the evidence being used to claim success is much stronger than it is currently, they are well positioned to operate effectively in a more connected environment.
Chapter One: Introduction
CHAPTER 2
MACRO POLICY CONTEXTS

What constitutes policy in relation to school improvement is a complex matter. It can manifest itself formally in legislation and government approved regulations and it can just as easily appear informally within emails, telephone discussions and face-to-face conversations. Sometimes situations arise where there is no relevant policy. Two common alternatives to move forward in such instances are to defer to related policies or to develop policy as things unfold. Policy in relation to school improvement is particularly complex because it is a combination of all of these things. Some researchers note that in the process of developing schooling improvement policy, well-intended policy developers typically set demanding improvement goals and then expect practitioners to reach those goals even though they do not necessarily have the requisite skills or knowledge to do so (Cohen & Hill, 2001; Elmore, 2005). Elmore (2005) captures the policy-practice breakdown as a circular dilemma.

Teachers and administrators generally do what they know how to do — they do not deliberately engage in actions they know will produce substandard performance, nor do they intentionally withhold knowledge that they know might be useful to student learning. If schools are not meeting expectations for student learning, it is largely because they do not know what to do. And, given the long standing disconnect between policy and practice, neither do policymakers. In its least desirable face, educational reform can become a kind of conspiracy of ignorance: policymakers mandating results they do not themselves know how to achieve, and educators pretending they do know what to do but revealing through their actions that they don’t. (p.217)

This policy-practice predicament is part of the learning dilemma. The contents of this chapter help to explain at the macro policy level of operation how governments in England, the United States and New Zealand are trying to work through the predicament. The upcoming investigations in Parts II and III help to explain at the micro level how participants within the seven school improvement initiatives selected for analysis are dealing with it.
Chapter Two: *Macro-Policy Contexts*

The focus on the macro policy context in this chapter is not intended to disregard the critical flow of influence by schooling improvement participants on government. International and New Zealand researchers assert that government disregard of participants can alienate and underutilise the knowledge and skills of those very people whose contributions are vital to successful reform (Elmore, 2005; A. Hargreaves & Fullan, 1998; Kliebard, 2002; Wylie, 1999). This is where the interaction between the research and development community and the practice community comes into play. Those two communities are constantly being influenced by government policies and, at the same time, they are constantly influencing policy directions. Ascertaining the extent of government influence on the research and development community and the practice community, and vice versa, is an important theme in the theory development in this thesis. Hence, the focus here on macro-policy contexts.

Before describing the macro-policy contexts specific to seven school improvement initiatives in England, the United States and New Zealand, it is useful to explain a broader global reform that influenced schooling developments in most OECD countries. This reform concerned developed countries’ governments thinking about the way their bureaucracies should manage public affairs (Kettl, 2000; New Zealand Treasury, 1987). Downward trending economies in many developed countries, including England, the United States and New Zealand, in the late 1970’s and 1980’s caused governments to think carefully about the way they allocated public funds and took responsibility for solving social problems. They were weighed down by cumbersome and unpopular bureaucracies that were maintaining, rather than solving, social problems. The government planned to devolve considerable responsibility for service provision so that people in communities would take more control of their own lives and their life choices. Treasury’s (1987) advice to the New Zealand government indicated that systems that reflected this perspective could help engage disadvantaged students in schooling: “A system which enables choices and control by parents and families is more likely to lead to positive identification with the education process, particularly by disadvantaged individuals and groups, and is more likely to achieve equity and efficiency goals” (p.147). This perspective was intended to foster greater local ownership. In relation to the model in Figure 2 (p.24), showing the relationship between the macro policy context and school improvement the new thinking meant
that responsibility for schooling improvement was devolved down the vertical arrow to locals.

In pursuit of more efficient and democratic social policy, most countries adopted the principles of a new approach to managing public affairs called “new public management”. A key principle of new public management was to devolve ownership and service provision to local units of operation (Kettl, 2000). In terms of school governance, devolution translated into far greater responsibilities for school districts in the United States, local education authorities in England and individual schools in New Zealand. Part of increased local responsibility involved those units coming to grips with the use of business practices and accountabilities alongside their educational functions. One benefit for using the principle of local ownership was that central government agencies were able to concentrate more on regulating the efficient and effective use of public funds, rather than an actual service delivery.

Developed countries’ governments typically went about introducing the principles of new public management by altering their bureaucracies and macro policy frameworks. Two reform approaches emerged to make the changes (Kettl, 2000). One approach to change was making swift and wholesale changes. The other was about reorganising things iteratively over time. In relation to the schooling improvement initiatives analysed in this thesis, England used a swift and wholesale approach, the United States used an iterative approach over time and New Zealand used a combination of both. The following explanations of the macro policy context in those three countries show how the three different approaches have been played out. Because this thesis primarily aims at helping improve the effectiveness of New Zealand’s school improvement movement, the description of New Zealand’s macro policy context is more detailed than that provided for the other two countries.

2.1. The Macro Policy Context in England

In England, schooling improvement initiatives such as the national literacy and numeracy strategies have been developed and implemented within a macro policy context of tight central control. The incoming Labour government in 1997 was intent on improving the quality of education system. Prime Minister at the time, Tony Blair, announced that education was the highest priority for his government. He said they aimed at “a world class education system in which education is not the privilege of
Chapter Two: Macro-Policy Contexts

the few but the right of many” (Blair in Earl et al., 2000, p.11). The government used its powers to take central control for improving performance in literacy and numeracy among all the students attending the 22,000 primary schools across the country. In the first instance, taking control meant setting challenging targets in response to the low overall performance of 11-year-old students in those two core curriculum areas5 (Department of Education and Skills, 1998). The targets anticipated that the percentage of 11-year-old students reaching level four in English would rise to 80 percent and in mathematics to 75 percent by 2002. Infrastructural supports put in place to help achieve the targets included redefining the national curriculum and unifying the bureaucracy to help schools deliver the curriculum effectively. Additionally, an outcomes-based accountability system was set up through a robust programme of standardised testing and “league table” publications of the results. Those alignments created a tight policy framework within which the literacy and numeracy strategies were expected to be developed and implemented.

The tightness did not end there. Monitoring mechanisms added to the tightness. There were; new national literacy and numeracy centres monitoring the introduction and use of curriculum frameworks and resources, a national standards and effectiveness unit within the Department of Education and Skills monitoring progress towards the challenging targets, an existing inspection unit (OFSTED) inspecting schools with a close eye on programme integrity; and, an international research team providing the national units with formative feedback to ensure the large scale approach worked (Earl et al., 2003). Those national monitoring mechanisms were linked to Local Education Authorities that helped practitioners in all primary schools across the country to put the national frameworks and resources into practice. Assisting the local authorities were 300 consultants spread across the country to ensure a high degree of programme integrity.

There appeared to be at least two reasons the English government chose to use tight central control over the reforms ahead of other policy possibilities. One reason cited by Michael Barber, a senior bureaucrat driving the reform effort, was to create the necessary levels of pressure and support to achieve the ambitious aims (Barber, 2000b). The external evaluation team concurred with his view and believed that the

5 A literacy taskforce set up in 1996 under the previous government found that only 65 percent of 11-year-old students were reaching the expected national standard for their age in English literacy (level four). The same exercise in mathematics completed by a second taskforce in 1997 revealed that only 61 percent of 11-year-olds students were reaching the expected standard for their age (level four).
English reform leaders succeeded in applying the necessary levels of pressure and support to succeed (Earl et al., 2003). For instance, on the one hand, pressure was being applied through the various evaluation agencies within the education system\(^6\). On the other hand, strong support was provided through high quality professional development and paper and electronic resources. The other reason was a political imperative to address entitlement issues for students in disadvantaged communities. Urgent attention to longstanding neglect of students living with disadvantage was an important election theme for the Labour Government throughout the election campaign. Challenging targets and “back to basics” practices, which could be applied quickly, engendered the sort of urgency for which the government was looking.

Some researchers argued that the tight central control and back to basics practices were more of a political imperative to impress voters for the next election rather than a genuine attempt to address long-term neglect (Brown, Millet, Bibby, & Johnson, 2000). I argue that the government was trying to pursue both political imperatives at the same time. Political popularity levers to gain votes and making evaluative information about schooling more available to the public are integral to modern day politics in developed countries. In this instance, Prime Minister Blair unashamedly took politics into education. However, the international research team informing the reforms indicated that the tactic did not appear to detract from the overall positive results or from leaving the door open to alter things if evidence suggested that was necessary (Earl et al., 2003).

The explicit and intentional action from the centre made clear distinctions between the responsibilities of government agencies, the schools and the research and development community. Michael Barber explained this intentionality by saying “Those responsible for reforming public education are in no position to deal with uncertainties. What they can do is manage and transfer knowledge about what works effectively, intervene in cases of under-performance, create the capacity for change and ensure it is flexible enough to learn constantly and implement effectively” (Barber in Earl, Fullan, Leithwood, & Watson, 2002, p.10). The need for clarity and direction did not allow the complexities, which are typically associated with school improvement, to cloud role definition. The final report of implementation

---

\(^6\) Evaluation agencies involved in the strategies included: the Office for Standards in Education (Ofsted) and the Standards and Effectiveness Unit (SEU) within the Department of Education and Employment which also commissioned the external evaluation team from the Ontario Institute for Studies in Education of the University of Toronto.
recommendations made it patently clear who was expected to do what (Department of Education and Skills, 1998). Government agencies, for instance, were left in little doubt about what to do and when to do it. An example of the clear expectations was in the district-level planning directions which stated that, “by April 1999 all LEA’s should submit Education Development Plans (EDP) on a three-year rolling basis detailing how standards will be raised in schools in their area. EDP’s should give high priority to literacy” (Department of Education and Skills, 1998, p.3). The same sort of clear direction was given to all others involved in the nationwide reform including the 300 consultants helping the 190,000 practitioners in 22,000 schools make sense of their tasks.

To summarise, the macro-policy context set up by the government in England left little doubt for professional educators about what and how the literacy and numeracy reforms were going to be developed, implemented and evaluated. It demanded a swift and wholesale change approach. The improvement agenda was set up within a few months after the election to urgently address a long-standing neglect of students within disadvantaged communities. Government’s pro-active stance on improving the education system as a priority task paved the way for its Department of Education to apply considerable pressure and support on schools from the centre by setting challenging achievement targets, unifying the bureaucracy and making everyone including the public aware of progress.

2.2. The Macro Policy Context in the United States

In contrast to the tight central control in England, the macro-policy context in the United States afforded schools and school districts a great deal more autonomy in the way schooling improvement initiatives could be developed and implemented, although the national standards were just as challenging as the targets set in England. The standards in this case centred on assuring that all students would be 100 percent proficient in reading and math by 2013-2014 (United States Department of Education, 2006). In a letter to chief superintendents of each of the 52 states in 2002, the Education Secretary outlined details of the obligation, which focused on submitting an accountability plan to the Department of Education (Paige, 2002). The plan had to explain how they were going to move towards achieving the standards and what they were going to do in the event that some school districts or schools did not reach these
Chapter Two: Macro-Policy Contexts

standards. An important part of the plan in this regard was to define “adequate yearly progress” to determine the achievements of each school district and school in order to highlight the improvement needs and to allocate adequate resources to address the needs. Upon receipt, the plans went to a peer review panel which made comment on how well the requirements of the statute were met. The peer review process for the Department was intended to be just as rigorous as the states’ planning requirements. Together, the state-wide planning and central review processes established a challenging standards-based school improvement environment.

The challenging standards did not simply appear overnight in the United States as did the targets in England. Rather, they were developed iteratively over many years. The iterative process manifested itself through an incremental tightening of national achievement standards over the last two decades. A useful way to explain the tightening process is to track federal government policy for distributing Title I funding, which is additional funding to help schools address issues related to disadvantage. A policy review in 1986 called ‘A Nation At Risk’ questioned the usefulness of remedial programmes to do the job and from then until the present day there have been four major policy revisions. Those revisions altered policy thinking around the use of Title 1 funding from a high tolerance towards remedial programmes

7 Ten criteria are used to assess the extent to which each state has met requirements. The 10 criteria are (1) A single state-wide accountability system applied to all public schools and LEAs; (2) All public school students are included in State accountability system; (3) A state’s definition of ‘adequate yearly progress’ is based on expectations for growth in student achievement that is continuous and substantial, such that all students are proficient in reading and math no later than 2012-13; (4) A State makes annual decisions about the achievement of all public schools and LEAs; (5) All public schools and LEAs are held accountable for the achievement of individual subgroups; (6) A State’s definition of ‘adequate yearly progress’ is based primarily on the State’s academic assessments; (7) A State’s definition of ‘annual yearly progress’ includes graduation rates for high schools and an additional indicator selected by the State for middle and elementary schools (such as attendance rates); (8) Adequate yearly progress is based on separate reading/language arts and math achievement objectives; (9) A State’s accountability system is statistically valid and reliable; and, (10) In order for a school to make ‘annual yearly progress’, a State ensures that it assessed at least 95% of students in each subgroup enrolled” (Paige, 2002, p.2 & 3).

8 Four key legislative revisions for Title 1 funding have been conducted since 1986. (i) A policy review entitled ‘A Nation At Risk’ in 1986 that questioned the usefulness of the funding to support remedial programmes led to a 1987 amendment which gave schools with 75 percent or more at risk students the option of using the funding for school-wide reform. (ii) A second policy review in 1993 found that the school-wide reforms were having a greater impact on academic achievement among disadvantaged students than remedial and self-managed strategies (Borman et al., 2003). A subsequent amendment encouraged schools with 50 percent or more disadvantage students to use the funding for school-wide reform. (iii) A third amendment in 1997 known as the Comprehensive School Reform Demonstration introduced nine criteria for schools to access the funding. The nine criteria matched the characteristics of the most successful school-wide reforms at that time, such as Success For All; measurable goals; support from staff members; research-based methods; external assistance; parental community involvement; staff development; co-ordination of resources; evaluation; and comprehensive school-wide reform plan. (iv) Two additional criteria were added to the original nine in a fourth
to a demand for comprehensive school-wide reforms and scientific evaluations of their effectiveness.

In contrast to the overt centralisation used in England for working out how to meet the challenging national standards, devolution was used as a key mechanism in the United States. Two devolved forms of school improvement emerged. One form was Comprehensive School Reform programmes. The federal government acted as an accreditation agency to distribute Title 1 funding to school districts and, in some instances, individual schools to implement Comprehensive School Reforms. In doing so, the federal government developed a direct contractual relationship with school districts and schools and an indirect contractual relationship with the researcher-developers of the programmes. School districts only got the additional funding if the programmes they selected met the federal government's stringent criteria. So there was an element of central control, albeit by contract. Most school districts developed contractual relationships with the research and development community to work out what to do. That was because developers within the research and development community had designed programmes that met the criteria. Consequently, although devolution afforded school districts considerable autonomy, most of them used their autonomy to conform to a programme designed by someone in the research and development community.

As national standards tightened, the developers of Comprehensive School Reforms such as Slavin who created Success For All, and the schools involved in the programmes, have had to change how the programme is designed, delivered and evaluated. Success For All was one comprehensive school reform programme which led the way in this regard. A recent study about the programme's impact on students' reading ability in 41 schools, for instance, advanced the quality of the methodology being used to show success from the use of quasi-experimental designs to a randomised design (Borman et al., 2005). That step up has met policy expectations but the researchers indicated in their reflective comments that they were intent on going one step further. "In future research, we will examine how well the programme theory matches the outcomes observed in the various reading assessments" (p.18). This intention means that the researchers are not only interested in producing

amendment in 2001 introduced through the No Child Left Behind Act. They made a comprehensive approach and scientific evaluations requirements rather than options. Scientific meant measuring improvement through third party experimental and quasi-experimental designs.
evidence of something positive happening but they are also keen to reveal causal relationships between the design theory of the programme and the results.

A second form of devolved school improvement is district-level reform, which also used the principle of control by contract. The way the Department of Education’s district office in New York District #2 operated throughout most of the 1990’s is an example of the contractual relationship between local government officials and school leaders. District officials negotiated with principals how they intended achieving the national standards (Elmore & Burney, 1997). Negotiations between the two parties were informed by a research team as well as by a district office official appointed specifically to help schools make links between standards-based policy and pedagogical practices (Elmore & Burney, 1998, 2000). What the negotiation process amounted to was a need to adjust down achievement standards for schools with hard-to-teach conditions. Although that decision sent an implicit message back to government that their national standards were important but not entirely realistic, it did not stop outcomes-focused accountabilities coming into play. Researchers found that in many instances the consequences for not measuring up to the superintendent’s expectations were serious (Elmore & Burney, 1998). In the decade when achievement results rose markedly in New York District #2, two thirds of the principals and over half the teachers were replaced.

The iterative reform approach in the United States did not create the clear boundaries between the policy, practice and research and development communities that characterised England’s swift and wholesale approach. In terms of monitoring schooling improvement initiatives, for instance, devolution blurred the boundaries between the policy and the research and development communities in the way that task could be completed. A variety of monitoring approaches were taken by the 52 state offices and their district offices. Some kept tight control of what was going on while others allowed for considerable autonomy. Researchers tracking improvements in New York District #2 had high praise for the quality of the tight monitoring by the officials in that district but suggested that their approach was rare (Elmore & Burney, 1998). If that is true then it would appear that the monitoring of school improvement initiatives by government officials at a local level in the United States was variable compared with that in England. Feedback from one of New Zealand’s senior policy analysts who meets regularly with visiting educators and officials from the United States confirmed the variability,
Many of the reform ideas in the United States such as the Comprehensive School Reform projects are all essentially voluntary activities on behalf of school districts. Many school districts can (and they do) invent their own curriculum and teacher hiring rules and there is nothing that a central government can do about it. I just think the United States system is complex and there is no doubt in my mind that some school districts in the United States are very jealous of other districts’ autonomy. (National senior policy analyst, New Zealand Policy Developers’ Feedback, 2005)

To conclude, challenging national achievement standards served as policy directives that state-wide and district-level officials, schools and the research and development communities were expected to understand and achieve. Many different intervention and monitoring arrangements eventuated. In the case of the comprehensive school reforms, federal government officials nurtured the research and development community into working things out with the schools. By contrast, in New York District #2, the government officials took an active role in designing the interventions and then monitoring the way the interventions were implemented. Overall, the various arrangements represent a complex matrix of collaborations trying to respond to the challenging national achievement standards.

2.3. The macro policy context in New Zealand

The macro policy context influencing the development and implementation of schooling improvement initiatives in New Zealand is a mix of central control and local autonomy. The mix came out of a swift and wholesale reform of education administration in 1989 followed by an iterative development of an intervention framework. It was out of these two developments that schooling improvement initiatives evolved.
2.3.1. **Influence of the 1989 administrative reforms**

The radical revision of the Education Act in 1989, commonly referred to as “Tomorrow's Schools” was a swift and comprehensive way of reforming administrative arrangements for schooling (New Zealand Government, 1989). Central to the administrative reforms was the principle that schools should govern and manage their own affairs (Picot Taskforce, 1988). Schooling had been administered by an overly bureaucratic Department of Education which had assumed responsibility for almost every decision there was to be made by compiling 45 volumes of regulations. The self-management legislation led to the transformation of the department into a lean policy-focused Ministry of Education. It also led to the devolution of considerable governance and management responsibilities to schools. A clear distinction between policy and practice set the scene for the Ministry's policy developers to distance themselves from the practice community and leave them to work things out for themselves.

Each New Zealand school is governed by its own parent elected Board of Trustees. A board of trustees is generally made up of the principal, five or more community members, and a staff representative and, in secondary schools, a student representative. These boards became critical to schooling improvement initiatives operating within New Zealand schools because of their legal powers. Under the “Tomorrow's Schools” legislation, they assumed responsibility for administering school finances, hiring the principal and staff, maintaining the buildings, consulting with the local community and, of most relevance to this thesis, making sure that the national curriculum was implemented. The boards are not told what they should emphasise in the national curriculum or how to go about ensuring it is implemented effectively. Rather, there are a plethora of guidelines available to help them work those things out for themselves.

---

9 The national curriculum, referred to in official documents as the New Zealand Curriculum Framework, is the official policy for teaching, learning and assessment (Ministry of Education, 2006). It has four components. (i) A set of principles that situate the individual student as the centre of all teaching and learning, that demand high quality curriculum and that allow for local discretion. (ii) Seven learning areas essential for a broad and balanced education including; language and languages, mathematics, science, technology, social sciences, the arts and health and physical well-being. (iii) Essential skills to be developed by students including; communication, numeracy, information, problem-solving, self management and competitive skills, social and co-operative, physical skills and work and study skills. (iv) The place of attitudes and values in the school curriculum.
Guidelines for curriculum implementation come in a variety of forms. There are a set of national education and administration guidelines\(^\text{10}\) to help boards of trustees work out governance and management systems and priorities. Those guidelines give boards flexibility to prioritise some learning areas over others depending on the academic and development needs of the student population. There are also numerous published frameworks, booklets and resources to help guide boards and practitioners to implement the curriculum and monitor its implementation. The Ministry has set up a web-based online learning centre which brings together many of the guidelines for practitioners, boards of trustees and the wider education community (Ministry of Education, 2006j). Practitioners, for instance, can go into a professional learning section (http://www.tki.org.nz/r/governance/prof_learn/) and browse through a range of professional learning opportunities available to them to develop their knowledge and skills in leadership and curriculum content. Each page typically provides a brief overview of the nature of the opportunities available, the emphasis for the current year and contacts for getting involved and for further information. All the guiding information from that website as well as from numerous other sources seems to mould into a matrix of guidelines from which boards and practitioners can pick and choose the bits they prefer.

Even though some researchers believe that the extent of the boards’ local control was both considerable and radical (Fisk & Ladd, 2000), others have noted that the boards were not let loose to act as they pleased (Gordon & Codd, 1990; Wylie, 1997, 1999). The latter researchers point out that the national guidelines constrained boards to comply with the centralised curriculum agenda. Additionally, several reviews and revisions of the national guidelines and an additional piece of legislation called the Education Standards Act (New Zealand Government, 2001b) tightened the policy boundaries and shifting the overall reform focus from improving administrative systems to improving educational aspects of schooling. The combination of the

\(^{10}\) National Education Guidelines have four components. (i) National Education Goals focused on statements of desirable achievement and government policy objectives. (ii) Foundation curriculum policy statements about teaching, learning, and assessment. (iii) National curriculum statements about areas of knowledge, understanding and skills to be covered by students and desirable levels in those areas. (iv) Six national administration guidelines intended as desirable principles of administrative conduct. They are; to foster student achievement through teaching and learning programmes which are consistent with the New Zealand Curriculum, to develop personnel policy and act as a good employer, to administer the finances and property, to provide a safe physical and emotional environment and, to comply with all general education legislation.
revised guidelines and the new legislation advanced the standard planning and reporting tool known as the "charter" from a values-based visioning document to include detailed strategic and annual plans that set goals and explain how they will be achieved. The guidelines also expected boards to review their plans and report to their communities and to the Ministry any differences between the goals set and actual achievements.

It is within that planning and reporting framework that boards serving disadvantaged communities decide what gets on the improvement agenda, what targets they will aim for and how they will go about achieving the targets they set. As a general rule, there is considerable flexibility in all three of these tasks. That means that they can choose to participate in whatever strategies and initiatives are on offer from the Ministry, education trusts and consultants or they can go develop their own.

In relation to setting targets, the national achievement standards are not as definitive or as challenging as those set in England or in the United States. Achievement standards in the national curriculum are expressed as levels with generous timeframes for students to move from one level to the next (Ministry of Education, 2006g). In addition there are no national tests until the final three years of secondary schooling. A range of different assessment tools have been developed\(^\text{11}\) to help schools measure students' movement through the curriculum levels. The use of these tools is voluntary. Underlying the voluntarism is a preference in the national assessment policy for teachers to work out with their students what achievement is all about, "Teachers and students need to develop a shared understanding about what is meant by achievement, what progress means, and what constitutes quality work" (Ministry of Education, 2001, p.2). The preference means that evaluation is first and foremost a formative task in the classroom. This is in stark contrast to the links between summative evaluation and external accountabilities in England and the United States.

\(^{11}\) Assessment tools recognised in national assessment policy to encourage greater understanding between teachers and students include: Diagnostic survey at entry to school (School Entry Assessment), online assessment of students' levels (National Curriculum levels 2-5) in mathematics, science and English (Assessment Resource Banks), tasks released by National Education Monitoring Project, benchmarking students' levels of performance against nationally moderated examples of student work (Exemplars), an electronic package of normative assessment tools for teaching and learning (asTTLe), the Supplementary Test of Achievement in Reading (STAR), the Performance Achievement Tests (PATs) and the Essential Skills tests.
Chapter Two: *Macro-Policy Contexts*

As a consequence of the high levels of devolution, flexibility and voluntarism, deciding whether schooling improvement is necessary and how it is accomplished is very much in the hands of boards of trustees and practitioners. A policy assumption underpinning the plethora of guidelines is that school boards and practitioners know what to absorb and what to set aside. In other words, there is considerable systemic confidence in their capacity to choose wisely. The high level of confidence, in turn, assumes that school boards and practitioners are able to make sense of the policies and ideas. It is also assumed that school leaders supervise and monitor classroom teaching. Those policy assumptions may be relevant and appropriate for schools serving student populations that are succeeding in school but they are not useful as generic principles for schools serving disadvantaged students (Annan, Fa'amoe-Timoteo, Carpenter, Hucker, & Warren, 2004).

2.3.2. Influence of the intervention framework

The argument developed in the previous paragraph suggests a need to examine the extent to which school boards serving disadvantaged communities are able to voluntarily select and use national guidelines and research ideas relevant to solving underachievement problems. Such an examination has been underway for some time now. The full extent of the flexibility and voluntarism only applies to adequately performing schools. An intervention framework allows the Ministers of Education and the Secretary for Education to intervene if they deem it necessary. What I mean by "if they deem it necessary" is that government’s intent was not to intervene if at all possible, hence a minimalist approach. Government asked the Ministry to develop the framework to help those boards that were struggling to self manage after the 1989 reforms and to protect the interests of students in struggling schools. The framework evolved into six intervention options12, which range in severity from a request for information through to the replacement of a board with a commissioner.

---

12 The only type of intervention built into the 1989 legislation was for seriously inadequate performance in governance (New Zealand Government, 1989, s106-109). The Minister of Education could dissolve a board of trustees and appoint a commissioner. That intervention was useful for extreme governance problems but it did not cater for a range of other less complicated situations. Alternative intervention support for boards experiencing financial difficulties came in the form of financial managers in 1992 (New Zealand Government, 1992, 81b) by specialist advisors in 1998 to help deal with other difficulties that did not warrant dismissing the board (New Zealand Government, 1999, 64A). Then in 2001 another amendment to the Education Act pulled together the three existing intervention options and added three more to provide a comprehensive range of supports (New Zealand Government, 2001, 1999).
Government’s request for an intervention framework and its evolutionary development was a response to demands for help surrounding approximately five percent of schools continuing to experience difficulties with self-managing their affairs after the Tomorrow’s Schools legislation took effect (McCaughey & Roddick, 2001). Three sources were found to create the demand for help. One source was principals indicating that self-management brought with it considerable workload demands and inadequate resources (Principals’ Implementation Task Force, 1990; Wylie, 1991, 1993, 1994). A second source was the Ministry and the government agency responsible for evaluating the effectiveness of schooling, known as the Education Review Office (ERO). The third source was groups of schools that self-identified a need for help. Three intervention approaches evolved concurrently to help schools deal with their administration and achievement problems within the macro policy context of devolution, voluntarism and minimal intervention.

**The first approach** was to support individual schools to improve their performance in administrative matters. Schools that the Ministry found to be facing difficulties were evaluated against a set of risk indicators. The indicators gauged the extent to which a school was a viable business unit. Schools that were most at risk of not surviving in the self-managing environment were given support for the time it took to address the problems. Achievement problems were not necessarily ignored in this first approach. Rather, there appeared to be a considerable number of other problems that needed to be addressed in order for the struggling schools to survive. Typical survival issues were roll decline, financial mismanagement and dysfunctional governance and management relationships. In summary, the approach was a contractual relationship between the Ministry’s Schools Monitoring and Support division and school boards experiencing difficulties, which was intended to safeguard government policy that all New Zealand schools were self-managing. I call it a contractual relationship because

---

Government, 2001a, 78. They include: (i) the Secretary for Education requiring a board to provide specific information (78J); (ii) the Secretary for Education requiring a board to engage specialist support (78K); (iii) the Secretary for Education requiring a board to prepare and implement an action plan (78L); (iv) the Minister of Education removing a board’s specified powers and directing the Secretary to appoint a limited statutory manager while the board remains in existence (78M); (v) the Minister of Education dissolving the board and directing the Secretary to appoint a commissioner because of performance issues; and (vi) the Secretary of Education dissolving a board and appointing a commissioner because of viability issues.

13 The 17 risk indicators are: curriculum issues, health and safety, governance and management, community relationships, leadership and management, Maori and Pasifika students, roll decline, roll change differential, operating deficits, learning resources, working capital, beginning teachers, teacher turnover, principal turnover, suspensions, stand-downs, school leavers with no qualifications, National Certificate of Educational Achievement (NCEA) Level 1, NCEA Level 2 and NCEA Level 3.
the Ministry support was provided through a contract called a ‘funding provision agreement’, which includes specified goals, set tasks, funding limits and timeframes for completing the tasks (Ministry of Education, 2002a).

**The second intervention** approach was one used by the Ministry’s monitoring and support division\(^{14}\) to help groups of schools that had high ratings against the risk indicators and that had high numbers of disadvantaged students. The idea was for groups of schools that met both those criteria to work together with local officials to solve their problems. An underlying principle of the intervention was adherence to a minimalist intervention approach to safeguard the overarching self-managing legislation (McCauley & Roddick, 2001). Consequently, school boards and lead practitioners were expected to take primary responsibility for leading the group interventions.

One senior Ministry official likened the approach to helping schools customise their own solutions locally in much the same way that cottage industries operate (L. Whitney, personal communication, November 28, 2001). Business literature refers to cottage industry as small localised groups of people painstakingly hand crafting products which reflect their local setting (Chandy, Sivilotti, & Kiniry, 2004). Since the mid-1990’s, the monitoring and support division has sponsored over 20 groups of schools to form cottage industry school improvement initiatives (Sinclair, 1999b). The SEMO initiative is one of those 20 cottage industry initiatives and is the second New Zealand schooling improvement initiative selected for analysis in this thesis.

The cottage industry approach was primarily a response to a series of accountability reports published by ERO\(^{15}\). The reports were about schooling difficulties in several districts with large numbers of students from disadvantaged communities (Education Review Office, 1996, 1997, 1998). Evaluative information in the reports revealed that some of the self-managing problems were more entrenched than Government first realised. They also indicated that the achievement problems exposed through the national and international surveys needed urgent

---

\(^{14}\) The Ministry’s monitoring and support division is officially called “Schools Monitoring and Support”. The division has been through several name changes. It started as the “At-Risk Schools Programme” in 1994. It was renamed “Schools Support” in 1997 after the term ‘at risk’ was questioned by education sector leaders as a suitable term for labelling schools in need of support. The current name of Schools Monitoring and Support was introduced in 2000 when the Monitoring Project was added to the division.

\(^{15}\) Additional evidence of struggle came from situational analyses by the Ministry’s monitoring and support division and from several groups of schools self-identifying a need for additional support.
attention in those particular districts. In response, the Ministry's monitoring and support division began intervening into groups of schools to address specific problems. Researchers examining the reforms in the two districts of Mangere and Otara found that the efforts, along with additional government funding, typically led to the use of old ineffective solutions (Timperley & Robinson, 2000). For instance, three schools chose to introduce a social skills programme as a lever for academic improvement (Annan, 1999). Yet they assumed rather than checked that the programme did actually lead to academic improvement. Despite the ineffective No.8 wire approach, the monitoring and support division persevered with the principle of strong local control of the cottage industry initiatives.

The third intervention approach was for the Ministry's national office to tailor professional development programmes across the country to address specific achievement problems. Feedback from the national senior manager overseeing professional development programmes said the approach was intended to break the mould of the national professional development programmes providing general curriculum information to practitioners that did not help solve specific achievement problems (New Zealand Policy Developers' Feedback, 2005). A typical design pattern of such programmes was for policy developers in the Ministry's national curriculum division (16) to analyse international and national evaluative survey information (17) for trends in order to determine professional development priorities. Priorities were, in turn, transformed into programmes that were contracted out to support organisations and consultants and delivered to interested schools across the country. Monitoring focused on inputs and outputs and there were typically no evaluations of the extent to which the programmes impacted on the achievement problems which they were intended to address.

The change to tailored national professional development programmes meant that the support organisations and consultants had to help participants analyse and address context-specific achievement problems rather than assuming that their local context problems matched the trends seen in the national and international data. The change

---

16 The curriculum division is called "Tertiary, Curriculum, Teaching and Learning," which recognises the Ministry's growing role in vital elements surrounding curriculum development. The label 'curriculum division' is used to reduce jargon for the reader.

17 System-level assessment sampling tools include the National Education Monitoring Programme (NEMP) and international surveys such as The International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA) (Ministry of Education, 2001).
also meant that all national professional development programmes had to be evaluated in order to check their impact on student achievement (Ministry of Education, 2006c). Within that change process from generic information distribution to tailored learning programmes, the Numeracy Development Project emerged as a successful schooling improvement initiative. It fitted the definition of a schooling improvement initiative in this thesis because it targeted all students to improve their thinking and performance in solving number problems.

In summary, the macro policy context in New Zealand of devolution, voluntarism and minimal intervention influenced the development of three different types of interventions. They included interventions into individual schools, cottage industry school improvement initiatives in clusters of schools and tailored national professional development programmes. The SEMO project emerged out of the cottage industry approach and the Numeracy Development Strategy was a national professional development programme. The three different sorts of interventions created opportunities for policy developers to become directly involved in solving a range of problems including underachievement after almost a decade of working at a distance from researchers and practitioners. They also gave boards of trustees choice in how to address their improvement agenda items, if, in fact, they perceived they needed such an agenda. That policy principle meant that regardless of the type of school improvement initiative used, design integrity remained contingent on the boards of trustees’ decisions to participate partially or fully in the initiatives.

2.4. Summary and Conclusions

This chapter described the macro policy contexts relevant to the selected schooling improvement initiatives in England, the United States and New Zealand. Each policy context was different and limited the type of school improvement that was possible. England’s macro policy context of tight central control meant that schooling improvement was mandated whereas the United States’ policy context nurtured diverse school improvement strategies. New Zealand’s macro policy context was different again, in that it was a mix of central control of a national curriculum and considerable autonomy in how the curriculum is delivered. That mix produced a combination of national and local schooling improvement strategies. There was also variation in the way the different policy contexts set expectations for success.
Whereas the English and United States contexts set challenging national achievement standards, the New Zealand context did not. In the absence of such standards, leaders of school improvement initiatives in New Zealand could look to proxy standards such as achievement levels attached to the national curriculum.

The diverse policy contexts highlight that government influences are contextualised within the politics and education system of each country. That is not to say, however, that government influences are unaffected by what is happening in other countries. There is general accord among researchers that global influences now impact on the governance decisions of individual countries, which in turn, impact on systems in education (Kettl, 2000; Olssen, Codd, & O’Neil, 2004). Whilst a great deal can be said about global-national influences, the focus of this thesis is mainly on national-local government influences because that is the context within which schooling improvement initiatives are developed and implemented. In Part II, attention turns to outlining how the five international initiatives were selected for analysis. The introduction reveals the criteria used to evaluate the effectiveness of the international and national initiatives and the two subsequent chapters critically analyse the condition of the evidence attached to the initiatives.
PART II: Investigating Evidence of Effectiveness

PART II

Investigating Evidence of Effectiveness

Purpose

Part II collates evidence on effective international schooling improvement initiatives in order to answer the first research question: What are the most effective national and international examples of school improvement and what is the condition of the evidence base for making claims of effectiveness? The evidence of central interest in this chapter is that which tells us whether a school improvement initiative is successful or not, i.e. statistically significant changes in student achievement information. Interrogating the evidence base at the outset of the investigations in this thesis ensured that theoretical arguments developed in later chapters about useful development and implementation ideas are explicitly linked to gains in student achievement. Those links are not assumed.

Approach

The investigations in Part II were conducted in three stages (i) stage one involved reviewing recent international literature on school improvement initiatives to develop indicators of successful initiatives, (ii) stage two used the indicators to identify and select five international and two national initiatives that best exemplified these criteria of success and, (iii) stage three interrogated the condition of the evidence found in the search for effective international and national school improvement initiatives.

Stage one: development of evaluation criteria. The inquiry involved establishing criteria to evaluate and select a sample of successful schooling improvement
initiatives from developed countries. Criteria for what counts as positive and statistically significant are based on those developed by Borman, Hewes, Overman and Brown (2003). They completed a meta-analysis that reviewed studies associated with 29 Comprehensive School Reform strategies in the United States. They developed four categories based on three criteria for their review. The three criteria provided the idea of combining the research methods with the magnitude of improvement to identify successful schooling improvement initiatives. The first two criteria focused on the quality and quantity of the evidence and the third criteria demanded a check of whether the results were positive and statistically significant.

Adaptations had to be made to the categories and criteria in order for them to be useful in evaluating the effectiveness of reforms from different countries. Other developed countries, such as England and New Zealand, are not using "Comprehensive School Reform" or third-party control group studies to measure effectiveness as defined in the United States. They are pursuing their own versions of school reform and methodological approaches for measuring effectiveness. For instance, in England the reform efforts selected for analysis have been evaluated using longitudinal studies with large sample sizes (Earl et al., 2000; Earl et al., 2003). By contrast in New Zealand, many small initiatives have operated with their own versions of measuring effectiveness.

Two adaptations were made to Borman et al.'s (2003) criteria and categories to accommodate the diversity of approaches across different countries. The first adaptation was to reduce the number of categories down from four to three; strong

18 "[The reform programmes were] grouped into four categories:
1. strongest evidence of effectiveness;
2. highly promising evidence of effectiveness;
3. promising evidence of effectiveness; and
4. greatest need for additional research.

The four categories were established on the basis of a combination of three criteria:
1. Quality of the evidence: Does the Comprehensive School Reform model have research evidence from the highest-quality studies: control-group studies and third-party control group studies?
2. Quantity of the evidence: Does the Comprehensive School Reform model have a relatively large number of studies and observations from which one may generalise the findings to the population of schools in the United States that are likely to adopt and implement Comprehensive School Reform models? (For instance, we used 10 or more studies overall and five or more third-party control-group studies as the, arguably arbitrary, standards necessary to be in the top category).
3. Statistically significant and positive results: Does the evidence from control-group studies show that the effects of the reform on achievement are positive and statistically greater than zero?" (Borman et al., 2003, p.154).
evidence, promising evidence and in need of further research. Borman et al's (2003) review concentrated on categorising a large number of initiatives and discussing which categories they fitted into. This thesis, however, concentrates on finding and analysing the initiatives with the strongest evidence of effectiveness. There was no need to discriminate between the initiatives as much as was the case in the original study. Consequently, the middle two categories of 'highly promising' and 'promising' evidence of effectiveness were collapsed into one 'promising' category.

The second adaptation was to add an extremely large sample size as an alternative to the '10-study' rule-of-thumb to accommodate schooling improvement initiatives other than Comprehensive School Reforms in the United States. The '10-study' rule of thumb was designed specifically for evaluating Comprehensive School Reform programmes. That style of reform lends itself to many studies because it involves experts developing a programme and selling it to a large number of schools or school districts. At the development end, the developers need evidence of effectiveness to sell their particular programme to many different localities. At the implementation end, the school districts and individual schools interested in participating in a particular reform require evidence of its impact on student learning to secure federal state funding for their initiatives. Hence, many comprehensive school reform programmes have produced a number of developer-led and third party studies in a variety of local districts.

Some research associated with versions of schooling improvement other than Comprehensive School Reform programmes, such as district-wide and national reforms, can produce strong evidence of effectiveness without replicating the results in ten or more studies. Strong evidence can be presented in comprehensive single studies which evaluate most if not the entire student population. That sort of evidence, whilst not as reliable as that presented from replicated studies, allows district-wide and national reforms to be considered for the sample in this chapter. Including the most successful of those approaches allows a broader analysis than would otherwise be possible into what constitutes an effective schooling improvement initiative.

Table 2 presents the three categories of effectiveness and their characteristics that were used to select that sample of initiatives involved in this thesis. The three categories of effectiveness are; initiatives with strong evidence of effectiveness, initiatives with promising evidence of effectiveness and initiatives in need of additional research. The characteristics for each category are set out in two points
which combine the magnitude of effectiveness (point a) with the quality of the research findings for each category (point b). In point b, two methodological criteria were used to test the quality of the evidence. The first was the size of the sample. The thinking behind that criterion was that the closer the sample size gets to the target population the more reliable claims of effectiveness become. The second was the extent to which the evaluation was replicated in different contexts. In this instance,

Table 2. Three categories for Selecting Effective Schooling Improvement Initiatives.

<table>
<thead>
<tr>
<th>Category</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiatives with strong evidence of effectiveness</td>
<td>Initiatives where:</td>
</tr>
<tr>
<td></td>
<td>a. Positive and statistically significant achievement effects are reported in the evaluations</td>
</tr>
<tr>
<td></td>
<td>b. Research methods include</td>
</tr>
<tr>
<td></td>
<td>• an extremely large sample size (all or almost all the student population) or</td>
</tr>
<tr>
<td></td>
<td>• replicated in a number of contexts (10 or more studies, five of which involve comparison groups or third-party comparison designs) and are reasonably generalisable to the population</td>
</tr>
<tr>
<td>Initiatives with promising evidence of effectiveness</td>
<td>Initiatives where:</td>
</tr>
<tr>
<td></td>
<td>a. Positive and statistically significant achievement effects are reported in the evaluations</td>
</tr>
<tr>
<td></td>
<td>b. Research methods include</td>
</tr>
<tr>
<td></td>
<td>• a moderate to large sample size</td>
</tr>
<tr>
<td></td>
<td>• replicated in a number of contexts (2-9 studies with two being comparison groups or third-party comparison designs) and are reasonably generalisable to the population</td>
</tr>
<tr>
<td>Initiatives in need of additional research</td>
<td>Initiatives where:</td>
</tr>
<tr>
<td></td>
<td>a. Positive achievement effects reported in the study/ies that may or may not be statistically significant</td>
</tr>
<tr>
<td></td>
<td>b. Research methods include</td>
</tr>
<tr>
<td></td>
<td>• A small sample size</td>
</tr>
<tr>
<td></td>
<td>• Less than two studies that may or may not be comparison or third party comparison designs.</td>
</tr>
</tbody>
</table>

19 Comparison group studies are those that compare the impact of an intervention on groups of students in different contexts, such as comparing results of students in one group of schools with those in another similar group of schools. The term 'context' relates to schools as the unit of analysis. Contexts within an individual school, such as year levels of the student population, as different contexts was considered tenuous because it assumes that there are no school level effects. Third-party studies are those conducted by researchers who are independent of the development of the intervention.
claims of effectiveness become more reliable the more times the results are replicated. So, for an initiative to be judged to have strong evidence of effectiveness, evaluations must demonstrate that it had a positive and statistically significant effect on student achievement. The research methodology underpinning the claimed achievement gains must either use a student sample that includes most if not the entire target student population or have replicated the results at least 10 times in different contexts.

In summary, the criterion developed to select the sample of national and international initiatives was strong empirical evidence of effectiveness. Two characteristics underpin this criterion. The first is that the evidence is positive and statistically significant. The second is that the evidence comes from research that used an extremely large sample size or replicated the findings in a number of contexts through 10 or more studies. These characteristics capitalise on Borman et al’s (2003) idea of combining the magnitude of improvement with strong research methodology to determine which initiatives are successful.

Stage Two: Selection of initiatives for analysis. The second step in the inquiry was to use the criteria to select a sample of successful initiatives from current international and national literature. This is where two parallel investigations were conducted; one to select international initiatives (Chapter Three) and the other to select national initiatives (Chapter Four). With regard to Chapter Three, a search of the international literature revealed two meta-analyses of many initiatives (Borman et al., 2003; Potter & Reynolds, 2002) and several outcomes-focused studies of individual initiatives (Earl et al., 2003; Elmore & Burney, 1998; Harwell, D'Amico, Stein, & Gatti, 2000; Jacob, 2003). The two meta-analyses presented evaluations of 29 and five initiatives respectively. The Borman et al (2003) meta-analysis judged three of 29 Comprehensive School Reform initiatives to have positive and statistically significant effects on achievement. They were Direct Instruction, Success For All and The School Development Programme. The researchers’ judgement was based on the results of more than 10 studies for each of the three initiatives. In the Potter and Reynold’s (2002) meta-analysis, two of the five initiatives reviewed, High Schools That Work and Success For All, were reported to have impacted strongly on student achievement.

Four other studies claimed three additional initiatives had strong evidence of effectiveness: Elmore and Burney (1998) and Harwell, D’Amico, Stein and Gatti
PART II: Investigating Evidence of Effectiveness

(2000) report the positive impact on student achievement of district-wide reforms in New York District #2; a study by Jacob (2003) reported the positive impact of the schooling reforms across the city of Chicago; and a research team from the University of Toronto reported achievement gains amongst students from 20,000 primary schools as a result of England's national literacy and numeracy initiatives (Earl et al., 2003). Combined, the two meta-analyses and four individual studies presented seven initiatives to be considered for the sample; Direct Instruction, Success For All, The School Development Programme, High Schools That Work, New York District #2, the Chicago reforms and England's national Literacy and Numeracy initiatives.

Unlike the international search of the literature, the New Zealand-based search found no meta-analyses describing the magnitude of the impact of school improvement initiatives on student achievement in New Zealand. In the absence of any such meta-analyses, the selection process involved locating evaluative information specific to many individual initiatives and then eliminating those initiatives with insufficient evidence of effectiveness. Researcher-developer research reports and progress reports from reform leaders to the Ministry20 were the two primary sources of evaluative information. The search found reports about 30 initiatives in total (28 cottage industry initiatives and two national professional development strategies). However, evaluative information about only seven of those was worthy of further analysis in light of the criteria being used in the selection process (Bishop et al., 2003; Higgins, 2001, 2002; Higgins, Bonne, & Fraser, 2004; Higgins, Irwin, Thomas, Trinick, & Young-Loveridge, 2005; Lai et al., 2004; Lai et al., 2005; McNaughton et al., 2004; Ministry of Education, 2004e, 2005a, 2005c, 2005f; Phillips et al., 2001; Thomas & Tagg, 2005b; Timperley, 2002, 2004a; Timperley, Phillips et al., 2003; Trinick & Stephenson, 2005; Young-Loveridge, 2005). The seven initiatives included five cottage industry initiatives, namely Achievement in Multicultural High Schools (AIMHI), Performance Enhancement in North Waikato (PEN), Students On A Rise in Paeroa (SOAR), Strengthening

---

20 "Progress reports" refer to documents prepared by reform leaders and Ministry officials that account for the expenditure of additional funding. Three forms of progress reports include: (i) a report prepared by reform leaders for Ministry officials about the progress of an initiative in order to release additional funding; (ii) a funding memo prepared by a Ministry official for senior Ministry officials to approve the release of additional funding (iii) an overview document of all the progress reports for one initiative prepared by Ministry officials in order to respond to requests from politicians, researchers and other interested members of the public for information about the initiative.
Education in Mangere and Otara (SEMO) and Te Putahitanga Matauranga (TPM), and two national professional development programmes, the Numeracy Development Project and Te Kotahitanga.

A template evaluative table (Table 3) was developed to verify the magnitude of the achievement gains and the quality of the research in the international and national initiatives. The table was set up in three columns. In the first column a brief description of each initiative is presented. In the second column evidence of improved academic performance of the students is summarised. Various researchers used different methods of reporting the magnitude of student achievement gains, such as effect-sizes, percentage gains or a qualitative rating scale. It was not possible to change the results into one standard measure. Instead, results are presented as they were reported in the source reports and evaluations in the second column of respective tables (Tables 7 & 8) in Chapters Three and Four. The third column rates each initiative for its evidence of effectiveness using the three categories described earlier, i.e. strong evidence, promising evidence or in need of further research. Initiatives with the strongest ratings in both investigations formed the sample for analysis.

Table 3. Evaluative Table to Select Initiatives with the Strongest Evidence of Effectiveness.

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Evidence of gains in student achievement</th>
<th>Effectiveness rating</th>
</tr>
</thead>
</table>
| List name of the initiative, form, number and type of schools, geographic location, any additional distinctive features. | List evidence of student academic achievement as stated in source documentation | Rate as either  
- strong evidence of effectiveness  
- promising evidence of effectiveness  
- insufficient evidence |

Stage Three: Method used to analyse the condition of the evidence. The third stage of the inquiry involved critical analysis of the condition of the evidence base associated with initiatives. The analysis is presented in two parts. The first part of the analysis concentrates on the quality of the evidence base in terms of the methodology used to measure the gains in student achievement and the magnitude of the gains. “Quality” is determined by the three categories of evidence. High quality is
PART II: Investigating Evidence of Effectiveness

synonymous with strong evidence, medium quality equates to promising and low quality equates to in need of further research. The second part of the analysis centres on reasons for the quality of the evidence. A constraints analysis table is used to analyse and make explicit the linkages. The constraints analysis technique comes from an applied research method called problem-based methodology (Robinson, 1993; Robinson & Lai, 2006). The technique helps to make explicit the practices that are being used to solve a particular problem, the thinking behind the use of those particular practices and the consequences of using them. Robinson and Lai (2006) use a number of examples to show how problem-based methodology can help people understand what and why they are doing things and the impact of their actions. For instance, they say:

Teaching is a complex network of practices. In any one teaching day, a teacher will engage in thousands of different practices, with each practice solving a problem about what to do in a particular situation. For example, if a teacher asks a rude student to leave her class, that is her solution to her practical problem of how to manage his behaviour..... Note that [the teacher] could have solved the problem in a different way. [She] could have thought of the student as needy rather than rude, and talked to him about how he experienced the lesson. (Robinson & Lai, 2006, p.17)

Robinson and Lai (2006) go on to explain that problem-based methodology involves considering a range of possible solutions and that those identified might not be good ones. It is a matter of inquiring into why a particular solution is preferred ahead of other solutions. The inquiring exposes the relevant theories attached to the practices being used, which provides an opportunity to evaluate the reasoning for acting in certain ways.

With the three-step inquiry process in mind for both the international and national investigations, Chapter 3 now focuses attention directly on the selection of the international initiatives for analysis and the condition of the evidence of effectiveness attached to those initiatives. Chapter 4 does the same for the national initiatives.
CHAPTER 3

EFFECTIVE INTERNATIONAL SCHOOLING IMPROVEMENT

The seven successful schooling improvement initiatives found in the literature to be worthy of the selection process are outlined in Table 4: Direct Instruction, Success For All, The School Development Programme, High Reliability Schools, New York District #2 reforms, the Chicago reforms and England’s national literacy and numeracy initiatives. The first four initiatives meet the criteria for Comprehensive School Reforms and the other three are district-wide, city-wide and nation-wide reform strategies respectively.

Table 4. International Initiatives with the Strongest Evidence of Effectiveness.

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Evidence of achievement</th>
<th>Overall rating of effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Instruction: principals trained as instructional leaders and teachers used highly standardised lessons in reading, language and mathematics using set resources.</td>
<td>• Overall effect size of $d = .21$ with 95 percent confidence interval(^{21}) of $d = .17$ to $d = .25$ from 48 studies with student achievement information. The effect size from comparison and third-party comparison studies was lower at $d = .15$ (Borman et al., 2003). Borman et al (2003) stated, “of the 48 studies in our analysis, most involved district comparisons or quasi-experimental matched-group comparison designs. A small number relied on the less preferred one-group pre-post design. Two studies used an experimental design with random assignment to treatment and control groups. The developer generated fewer than 10% of the outcomes” (p.187). • Strong evidence (Potter &amp; Reynolds, 2002)</td>
<td>Strong evidence</td>
</tr>
</tbody>
</table>

\(^{21}\) Borman et al (2003) state that, “the confidence interval expresses the degree of accuracy of the effect size estimate and suggests a range of effects that are likely to be found in similar implementations and studies of the reform model” (161).
| Success For All; highly scripted teaching in reading and mathematics, grouping students into classrooms with similar levels of achievement, eight weekly assessments and a family support team. | • Overall effect size of $d = .18$ with 95 percent confidence interval of $d = .16$ to $d = .21$ from 42 studies with student achievement information. The effect size from comparison studies is $d = .18$ and third party comparison studies were much lower at $d = .08$ but still statistically significant (Borman et al., 2003). Borman et al (2003) state, “of the SFA [Success For All] studies, most involved quasi-experimental matched-group comparison designs, some as complex as matching individual students across matched comparison schools. Only a handful of studies relied on the less preferred one-group pre-post design. Slightly under half of the outcomes were reported by researchers other than the developers” (p.206).  
• strong evidence (Potter & Reynolds, 2002) | Strong evidence |
| The School Development Program; assessment of student's developmental needs and teams of adults constantly adjusting the school climate. | • Overall effect of $d = .15$ with a 95 confidence interval of $d = .10$ to $d = .20$ from 18 studies with student achievement information (Borman et al., 2003). Borman et al stated, “For our analysis, 72% of the data was based on quasi-experimental matched-group designs. The rest of the data were based on one-group pre-post designs. Researchers other than the developers provided a little more than a third of the outcome data” (p.204). They also stated that the effects drop considerably for comparison studies at $d = .05$ and third party comparison studies at $d = .11$. However, these smaller effects were considered positive and statistically significant. | Strong evidence |
| High Schools That Work (HSTW); 1987 – The largest and oldest of the Southern Regional Education Board’s school improvement initiatives for high school and middle grades leaders and teachers. More than 1,200 schools in 32 states are using the framework of HSTW Goals and Key Practices to raise student achievement. | • Overall effect size of $d = .30$ calculated from 45 studies with student achievement information (Borman et al., 2003). Borman et al (2003) stated, “Most of the studies employed the same type of analysis: a cohort design that tracked the progress of HSTW schools by examining the outcomes of successive senior classes from year to year” (p.182). There was no indication of the number of studies generated by the developers compared to third-party researchers. Additionally, one comparison study revealed a statistically significant negative effect size of $d = -.06$.  
• Strong evidence (Potter & Reynolds, 2002) | Promising evidence |
**Chapter Three: Effective International Schooling Improvement**

<table>
<thead>
<tr>
<th>New York District #2; a district-wide instructional improvement plan. Key features included negotiated achievement standards for each school, principal training as instructional leaders, professional development for teachers, close monitoring through school visits and analysis of schools' performance in city-wide tests.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- In 1987 New York District #2 was ranked 16th out of New York's 32 community districts for reading and mathematics standardised tests. By the mid-1990's the district advanced to become the second ranked district in the city (Elmore &amp; Burney, 1997)</td>
</tr>
<tr>
<td>- Between 1988 and 1998, the percent of students achieving at or above grade level in reading rose from 56% to 73% and from 66% to 82% in mathematics (Harwell et al., 2000)</td>
</tr>
<tr>
<td>Strong evidence</td>
</tr>
<tr>
<td>Chicago reforms; ended social promotion – students in grades 3, 6 and 8 required to meet minimum standards in reading and maths Iowa Test of Basic Skills (ITBS). Students that failed required to attend 6-week summer school (15 year olds to transition centres). Academic probation to hold teachers and schools accountable – threat over schools with fewer than 15% students scoring at or above national norms being reconstituted – with principals and teachers dismissed or reassigned.</td>
</tr>
<tr>
<td>- In mathematics .3 standard deviation higher than expected in 2000 – significantly higher than other Mid-West cities, such as Cincinnati, Gary, Indianapolis, Milwaukee and St Louis (Jacob, 2003)</td>
</tr>
<tr>
<td>- In reading .2 standard deviation higher than expected in 2000</td>
</tr>
<tr>
<td>- Possibility of test inflation with these scores from one test</td>
</tr>
<tr>
<td>Secondary school gains included:</td>
</tr>
<tr>
<td>- more students on track one year after elementary school</td>
</tr>
<tr>
<td>- more first-year students pass an honours course</td>
</tr>
<tr>
<td>- passing rates for algebra/geometry sequence rise</td>
</tr>
<tr>
<td>- more students complete college preparatory course load</td>
</tr>
<tr>
<td>- graduation rates for 19-year-olds increased slightly and drop out rates for 19-year-olds decreased slightly (Miller, Allensworth, &amp; Kochanek, 2002)</td>
</tr>
<tr>
<td>Promising evidence</td>
</tr>
<tr>
<td>English Literacy and Numeracy initiatives; national frameworks for teaching literacy and numeracy that created a high level of standardisation through scripted lessons and resources, public assessment information and school inspections.</td>
</tr>
<tr>
<td>- In 1997, 63% of children reached the expected level in English (Target 80%). Increased to 75% in 2002 (Target 80%) (p.3. Summary and main report)</td>
</tr>
<tr>
<td>- In 1997, 61% reached target in mathematics. Increased to 73% in 2002 (Target 75%)</td>
</tr>
<tr>
<td>- Earle et al (2003) claimed the gap has narrowed substantially between pupil results in the most and least successful schools and LEA's.</td>
</tr>
<tr>
<td>Strong evidence</td>
</tr>
</tbody>
</table>
3.1. Selection of international initiatives for further analysis

Selecting the sample of international initiatives for further analysis was relatively straightforward. As five of the seven initiatives in Table 4 were judged to have strong evidence of effectiveness, they became the sample. They were Direct Instruction, Success For All, The School Development Programme, New York District #2 and England’s national Literacy and Numeracy initiatives. Of the three Comprehensive School Reform programmes, Direct Instruction had the strongest evidence of effectiveness, with an effect size gain of $d = .21$. Success For All’s overall impact was slightly below Direct Instruction with an effect size of $d = .18$. Slightly lower again was the third Comprehensive School Reform programme ($d = .15$). Borman et al (2003) found that the impression of improvement from these three initiatives dropped when they analysed only studies with third party and comparison designs. Direct Instruction dropped to $d = .15$, Success for All went down to $d = .08$ and the School Development Programme dropped considerably to $d = .05$. The lower effect sizes did, however, remain positive and statistically significant. One reason proposed by Borman et al (2003) for the lower scores from third party studies was because the researchers were independent of the programmes and were able to be more objective than the researcher-developer who were developing the programmes and evaluating them at the same time.

The other two initiatives selected for the sample and listed in Table 4, New York District #2 and England’s national literacy and numeracy initiatives both met the criteria for strong evidence of effectiveness. In the case of the New York District #2 reforms, Elmore and Burney (1998) reported significant gains in reading and mathematics after several years of district-wide instructional improvement strategies. Between 1988 and 1998, the percent of students achieving at or above grade level rose from 56% to 73% in reading and from 66% to 82% in mathematics, representing increases of 17 percent and 16 percent respectively. Other researchers confirmed that the remarkable improvement was significant (Harwell et al., 2000). After being ranked sixteenth out of New York’s 32 community districts for reading and mathematics standardised test results in 1987, the district advanced to become the second ranked district in the city by the mid 1990’s. In an evaluation of the English Literacy and Numeracy strategies, Earl et al (2003) tracked the number of 11-year-
olds in England reaching the expected standard of Level Four in national assessments from 1997 through to 2002. The researchers found a 12 percent increase in English the same increase in mathematics, both of which were considered significant.

Studies associated with High Schools That Work and the Chicago reforms raised cautions that caused them to be excluded from the sample. The High Schools That Work initiative rated highly in two studies (Borman et al., 2003; Potter & Reynolds, 2002). However, further analysis in one of those studies (Borman et al., 2003) found the encouraging results were based entirely on pre-post evaluations performed by the developers. Questions were raised however, when the results from the pre-post evaluations were compared with comparison studies. One comparison study, for instance, revealed a statistically significant negative effect size of d = -.06. More research was considered necessary to categorise the initiative as having strong evidence of effectiveness.

With regard to the Chicago reforms, the standard deviation gains (.3 in mathematics and .2 in reading) reported in one study also appeared to represent strong evidence of effectiveness (Jacob, 2003). However, the researcher raised two concerns about the evidence which suggested it was promising rather than strong. The first concern was that test inflation could have occurred from the single test measurement. A common problem with single test measures is that teachers can teach to the test items to make it look as if their students are doing well. The second concern was that an item level analysis of mathematics data indicated gains were focused in particular areas and did not represent an overall gain in mathematics. Based on these cautions, the Chicago reforms were categorised promising rather than strong evidence of effectiveness.

In summary, the search for evidence of successful international school improvement initiatives located 32 initiatives in the evaluation literature which appeared worthy of consideration for further analysis. From the pool of 32 initiatives, seven had evidence that was sufficiently developed to analyse further to see if it was strong or promising. From that analysis, evidence associated with five initiatives was rated as strong and evidence for the other two was rated promising. Those ratings meant that five initiatives with strong evidence of effectiveness (Direct Instruction, Success For All, The School Development Programme, New York District #2 and England’s national Literacy and Numeracy initiatives) would be analysed in order to
discover development and implementation characteristics attached to them that might help accelerate the learning of effective school reform practices.

3.2. The condition of the evidence

Findings about the condition of the evidence are presented in two parts. The first part presents information about the quality of the evidence and the second part examines the reasons for the quality found.

3.2.1. The quality of the evidence

Evidence of effectiveness associated with the five initiatives selected for further analysis is high quality. This claim is based on the finding that robust methodologies were used and the achievement gains were both positive and statistically significant. At first glance, the achievement gains may have appeared disappointing. Borman et al’s (2003) effect size gains from third-party and comparison studies of .15 for Direct Instruction, .08 for Success For All and .05 and .11 for The School Development programme are a case in point. Those gains contribute to an overall gain of .11, which does not appear to be anything to get excited about. However, educators have yet to agree on the magnitude of quantitative measures to claim success. One research perspective considers that those sorts of gains are at the lower end of effect sizes associated with innovations in the field of social sciences (Lipsey & Wilson, 2001). From that perspective, a small effect size is at or below .2 and a large effect size is at or above .8. In keeping with this thinking, Borman et al’s (2003) overall effect size gain of .11 for the three most successful comprehensive reforms is disappointing.

An alternative view in the debate is that statistically significant effect size gains at or below .2 should not be trivialised. Borman et al (2003), for instance, did not consider the overall effect size of 0.11 for the three reforms to be unsuccessful. They argued that it was a successful result from several perspectives. From a statistical perspective, it is both positive and statistically significant. From a pragmatic perspective, it means that 55 percent of the schools involved in Comprehensive School Reform outperformed control schools. From a historical perspective this result was better than the $d = .10$ effect gains made from previous initiatives funded
from additional federal funding. Not only did the overall effect improve by $d = .01$, but the sophistication of the measures also improved. Earlier one-group pre-post designs which Borman et al (2003) claim tend to inflate effects positively were superseded by control group studies and third party control group studies.

The stance that I take in relation to the debate about the value of the small gains made by the five international initiatives under review is that they are worthwhile. Two factors additional to Borman et al’s (2003) key points support this stance. The first factor is that the five initiatives are complex attempts to manipulate several variables at the same time with large sample sizes in difficult circumstances. For instance, they all intervene into teaching and learning as well as into management systems and, in the case of The School Development Programme, into families. There is widespread recognition that trying to intervene into multiple variables at the same time is both complex and difficult (Fuhrman, 2002; Fullan, 2005a; Potter & Reynolds, 2002; Rowan et al., 2004; Stringfield, 2000). So those initiatives that get positive and significant gains, even if they are small gains, are worthy of analysis to find out how to speed up the learning process.

The second additional factor in support of valuing small achievements gains is that a small gain can represent a breakthrough in changed attitudes or increased skills and knowledge among the practitioners teaching disadvantaged students or the students themselves. A good example of this is the claim that one group of researchers made about the effect of the enthusiasm of teachers involved in early trials of the English numeracy strategy (Brown et al., 2000). They claimed that achievement gains made motivated teachers involved in early trials to continue using the practices. Those teachers apparently spread their enthusiasm through collegial networks before the reforms officially started, which ensured a relatively high level of teacher buy-in from the outset. General enthusiasm among teachers towards the reform practices offset other factors that could have derailed the developments, such as high levels of prescription, related work load issues and the exposure of teachers to criticism from league table publications. This example may be a one-off for large scale reforms but the principle of leveraging off small successes to get momentum is a change principle that supports the valuing of small achievement gains.

Coming up with an acceptable overall benchmark for what counts as effective beyond achieving statistical significance is fraught with difficulty. One researcher considers that for school interventions to be rated as successful they need to show an
effect size of 0.4 (Hattie, 1999). Although this benchmark is useful as an optimistic target for leaders of schooling improvement initiatives, it has been criticised because of apparent discrepancies in compiling masses of studies to generate the figure (Nash & Prochnow, 2004) and it does not solely represent interventions manipulating multiple variables. It takes into account numerous studies, including those that report interventions measuring single variables with small groups of students and schools.

What does seem worthwhile is making judgements about whether small statistically significant achievement gains are educationally significant or not given the particular circumstances in which the gains were made. Achieving statistical significance is critically important but not sufficient to make a positive difference in the everyday schooling lives of disadvantaged students. The ideal is to create gains that are both statistically and educationally significant. That is not an easy task.

Achieving statistical significance is challenging but relatively straightforward in that it demands rigorous measures of success and mathematical calculations to verify the accuracy of the measures. However, the outcomes-focused studies reviewed for the selection process in this chapter indicate that identifying the factors that are making an educational difference or that are stopping things from getting worse is extremely difficult. For instance, one of the studies which revealed the achievement gains in New York District #2 tried to pin point the difference that the professional development was making (Harwell et al., 2000). The research team could only claim that it was one factor alongside many others that appeared to be contributing to the success. They could not say what the exact mix was. That inconclusiveness is reflective of the international schooling improvement research community only being in the early stages of working out what actually makes the difference.

3.2.2. Reasons for the high quality evidence

The constraints analysis in Table 5 explains why the evidence of effectiveness attached to the five international initiatives was high quality. It outlines two priority practices and the constraints that led to those particular practices being used ahead of others. One practice is investing in summative programme evaluation mainly to satisfy the demands of external accountability systems. The other practice is the use of expert programme evaluators with the capability of producing high quality
evidence to conduct outcomes-focused evaluations. This practice is primarily the result of clearly defined roles for groups participating in schooling improvement initiatives. Each practice and its associated constraints are explained more fully in the sections immediately following Table 5.

Table 5. Constraints Analysis for the Strong International Evidence.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Practices</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>• External accountability</td>
<td>• Investment into summative programme evaluation</td>
<td>• High quality evidence of effectiveness</td>
</tr>
<tr>
<td>• Role clarity</td>
<td>• Use of expert programme evaluators to conduct evaluations</td>
<td>• Critical debate about the evidence within the research and development community</td>
</tr>
</tbody>
</table>

3.2.2.1 Investment into programme evaluation

Evidence that there was adequate investment in summative programme evaluation came in two forms. One form was the infrastructural supports that made it possible to produce high quality evidence of effectiveness. Programme evaluators had readily available achievement data stored in systemic achievement information management systems. For instance, the international researchers tracking England’s national literacy and numeracy strategies had access to achievement information on all primary school students stored in an electronic software package called “Pupil Achievement Tracker” (PAT) (Department of Education and Skills, 2006). Schools submit their data annually using a standardised form and once it is inputted into the system the national centre directors and their international research team, local
officials and school leaders can track trends for individual schools or groups of schools against national norms at school and individual student levels of analysis. Such systems require standardised common assessment tools, pooling of data and interrogation of the data to find within and across school achievement trends, and therefore require adequate investment into summative evaluation infrastructure.

The other form of evidence was the large scale nature of the outcomes-focused research activity. In the case of the three comprehensive school reforms, that meant lots of developer and third-party studies. Borman et al (2003), for instance, report 48, 42 and 38 studies attached respectively to Direct Instruction, Success For All and the School Development Project. In the case of the English initiatives, large-scale meant one major pre-post study that involved full coverage of all the students in the 22,000 primary schools across England (Department of Education and Skills, 2006). Full coverage was also accomplished by the researchers in New York District #2 tracking of all the students against the City-wide testing programme (Elmore & Burney, 1998).

Combined, I believe that the investments into infrastructural supports that made achievement information accessible to programme evaluators and the large scale nature the research activity helped to produce the high quality evidence attached to the five international initiatives.

There are at least three possible constraints that could explain why this high quality evidence was produced. The three constraints are explained in turn before outlining why one is favoured over the others as the most plausible explanation. Slavin and Comer’s involvement in evaluating Success For All and the School Development Programme respectively is used as an example to draw out the three constraints. The first constraint that led Slavin and Comer to invest in summative evaluations was their intrinsic interest in checking the usefulness of their theories and practical ideas. Both of these developers have spent many years researching and developing the ideas for their respective initiatives (Comer, 1996; Slavin, Madden, Karweit, Dolan, & Wasik, 1992).

The second constraint is the developers’ need to survive in the comprehensive school reform market place. Those two developers have never been guaranteed business in the market. They have to secure business from schools and school districts which typically search for a suitable programme to access Title I funding. There are a multitude of catalogues that have been developed for this purpose. A quick search on the world-wide net on Sunday 2 July 2006, for instance, found
1,960,000 results for ‘Comprehensive School Reform Catalogs’ and it took less than 30 seconds to locate evaluative information about Success For All and the School Development Programme. One catalogue which appealed for ease of use and for its inclusion of outcomes-focused research findings was the one produced by the Northwest Regional Educational Laboratory in Portland (Northwest Regional Educational Laboratory, 2006). It provides different drop down options to search for initiatives by student year groups or characteristics of various programmes. Once you find the sort of initiative you are looking for, explicit information is included about its success. In the case of Success for All, for instance, the introductory evaluative statement is: “Success for All, the reading/language arts component of Roots & Wings, has been evaluated extensively, with statistically significant positive results for program students compared to control students across many studies” (Northwest Regional Educational Laboratory, 2006). That sort of advertisement is critical to on-going business for the likes of Slavin and Comer. It is, therefore, important for them to invest in proving their success so they achieve high ratings in catalogues and other literature read by client schools and districts and others considering joining up.

The third constraint, which relates closely to the second one, is the need to satisfy the demand for robust scientific research in the federal government’s criteria for Title I funding (United States Department of Education, 2002). Without repeating details of those demands, which are outlined in the description of the macro policy context in the Unites States in Chapter Two (2.2), it is sufficient to say that comprehensive school reform developers have been increasingly challenged for over a decade to produce more accurate evidence verifying the value for money of their programmes. In summary, I am arguing that Slavin and Comer were probably investing in proving their success for career satisfaction, survival in the business and to satisfy external masters. I am also arguing that the developers of the other three initiatives were facing similar constraints in their investment decisions. For example in England, Barber and, more recently Hopkins are well established and respected educators who want to check that their decisions are sound, want to survive in their roles by proving the worth of the programmes to government and, want to show the public positive results so the government is held in a positive light.

Of the three constraints, the first is the least credible explanation for investing in evaluations of schooling improvement initiatives. It is probably fair to say that most if not all educators who have spent their careers in education are intrinsically driven to
further develop their theories and practices but not all invest in rigorous evaluations. It is my belief that the second and third constraints are the priority driving forces behind the investments into high quality evidence attached to the five international initiatives. I also believe that those two constraints are inextricably linked. Survival in leading international schooling improvement initiatives, whether that is in a commercial or political context, is contingent on results satisfying audiences beyond the developers' control. Satisfying schools participating in an initiative that things are going well is one thing. Satisfying government or the public or both has upped the accountability stakes considerably. Of the three constraints presented to explain the international initiatives' adequate investments into summative evaluations, I believe external accountability is the most influential followed closely by survival in the role, while altruistic feelings about doing some good underpin most activity in the education sector.

One last point about the links between investment practices and external accountabilities, which is listed in the constraints analysis in Table 5 as a second consequence of those practices, is the way they caused members of the research and development community to hold each other to account for the claims being made. Academic debates about effectiveness are common-place in the literature related to five international initiatives. There are claims and counter claims made of individual initiatives. For instance a rigorous debate was held in the late 1990's about whether the School Development Programme had a sufficiently strong instructional focus (Comer & Haynes, 1999; Cook et al., 1999). Cook and associates claimed from their investigation into the programme in 23 middle schools in Maryland that the programme had a social focus which was not driving up achievement:

> With the models used, we did not find that the programme improved the state math test scores. Indeed, the indications were that it might have caused them to drop relative to schools with a less Comer-like climate. If this finding were to be replicated in other Comer sites, it would suggest that the program's social emphasis can have negative side effects for achievement. (Cook et al., 1999, p.586 & 587)

Comer responded claiming that an academic focus was central to the Comer way of doing things and schools without such a focus were not adhering to the programme.
Chapter Three: *Effective International Schooling Improvement*

The authors also make several errors and omissions in their description of the School Development Project. First, it is not correct that the social climate focus in the School Development Project is more important than the academic focus. The Comprehensive School Plan, one of nine essential elements of the School Development Programme includes both a social and an academic focus. The absence of an academic emphasis at any programme site is an implementation problem. (Comer & Haynes, 1999, p.600)

There are also critical reviews that question results across initiatives. Borman et al’s meta-analysis of evaluations of comprehensive school reforms is a case in point. By gathering and analysing a considerable number of researcher-developer and third-party studies, the research team found a level of inflation in the results generated by the researcher-developers. The checks for inflation lowered the effect size gains for Direct Instruction from .21 down to .15, for Success For All from .18 down to .08 and for the School Development Programme from .15 down to .11. The key point here is that because of adequate investments into outcomes-focused research, the research and development community in the United States were able to debate the credibility of the evidence attached to schooling improvement initiatives.

3.2.2.2. Use of expert programme evaluators

The second practice of using expert programme evaluators to ascertain the level of success was apparent in all five initiatives. The practice manifested itself in several different ways. One manifestation was through two of the three comprehensive school reform developers, namely Slavin and Comer, taking it upon themselves to oversee evaluations of the effectiveness of their programmes. They are expert programme evaluators in their own right and have published considerably on the value of their own programmes (See Appendix B). Additionally, they now employ other expert evaluators within evaluation units in their respective headquarters in Baltimore and Yale University to assist them as the coverage of the programmes has expanded (Success For All, 2003; The School Development Programme Website, 2004).

Another manifestation of the practice was localised programme evaluators with high levels of expertise getting involved. This happened in Direct Instruction and
New York District #2. Siegfried Engelmann, the developer of Direct Instruction chose not to follow the same path as Slavin and Comer. He and his directors left it to the schools and school districts participating in the programme, with help from localised programme evaluators to evaluate the success of the programme. They made that decision based on belief that schools and school districts with the help of local researchers would produce the necessary evidence for their Title I funding applications (K. Engelmann, personal communication, January 8, 2004). It meant the directors were freed up to get on with developing the programme. Borman et al’s (2003) discovery of 48 studies investigating the effectiveness of Direct Instruction validates the directors’ decision and also suggests that capacity among localised programme evaluators in the United States is in good heart. The quality of the evaluations produced for the schools and officials in New York district #2 by nearby researchers (Elmore & Burney, 1998; Harwell et al., 2000) lends further support to the argument of the existence of expert programme evaluators at the local level. A third manifestation of the practice was the international research team contracted through the University of Toronto to evaluate the English Literacy and Numeracy strategies (Earl et al., 2002; Earl et al., 2000; Earl et al., 2003). That team brought with it considerable expertise and knowledge in programme evaluation.

Two constraints appeared to rule in the consistent use of expert programme evaluators to conduct outcomes-focused research across all five initiatives. One is the need to meet the demands of external accountability systems and the other is the pre-existence of role clarity. The argument in the previous section about the strong influence of external accountability systems on investment decisions (3.2.2.1) is just as relevant in this section. Role clarity is also a serious contender. What I mean by that is that the various groups involved in the international school improvement initiatives have clearly defined roles. Programme evaluation sat clearly in the hands of the national or international research and development community. It was not part of the work programme set for policy developers and practitioners. So evaluation was not only tied to expertise, it also had an element of independence. As it happened, the research and development community appears to have embraced the task and is measuring up to the evaluative challenges coming from the macro policy contexts. Slavin’s positive efforts to respond to demands for scientific research in the Title I funding release processes is a case in point (Borman et al., 2005).
Some of New Zealand’s senior policy developers expressed concerns about expert programme evaluators taking full responsibility for conducting summative evaluations of schooling improvement initiatives. A major concern was that the heavy reliance on the research and development community may cause practitioners to become passive evaluative thinkers. For instance, practitioners may be aware of particular achievement problems or problems with the way some programmes are trying to address identified problems. However, they are unlikely to question them if evaluating the programmes is not part of their role. It is fear that New Zealand educators will opt out of the evaluative component of teaching that limits policy developers commitment to external evaluation (Alton-Lee, 2006). However, safeguards can be built into reforms to stop that from happening. It is a matter of distinguishing between the programme evaluation function that the research community is leading and the formative assessment function that is the responsibility of practitioners in the practice community. There is also a need for strong feedback and feed forward learning loops between those two communities. The extent to which those links were in place in the five international reforms is part of the analysis in Part III.

3.2.3. **Summary of the condition of the evidence**

To summarise, the evidence attached to five international school improvement initiatives is high quality. Methodological rigour contributed to the high rating as did the small positive and significant achievements gains reported. Debate continues in the research community as to whether small gains such as the ones found should be discounted. I argue they should not, particularly when they are making an educational difference in students’ lives. Two practices were found to be associated with the high quality evidence of effectiveness. Investment into summative assessment was a taken-for-granted practice in the reform process. That investment did not discount the important place of formative assessment. Rather, summative assessments provided useful information for formative purposes in the design rooms of the initiatives as well as in the classrooms. The second reason for the high quality evidence was because expert programme evaluators conducted the evaluations. Demands from external accountability systems and clarity of roles were cited as the main constraints that ruled in those two particular practices and ruled others out.
3.3. Summary and Conclusions

The first purpose of this chapter was to identify a sample of successful international school reforms for further analysis in Part III. The criteria developed to complete this task found five international school improvement initiatives in the literature with strong evidence of effectiveness. That sample included three comprehensive school reforms (Direct Instruction, Success For All & the School Development Programme), a district-level approach (New York District #2) and a national approach (English literacy and numeracy initiatives). A combination of three different types of reforms provides an interesting mix of initiatives with strong evidence of effectiveness for analysing development and implementation characteristics in Section III.

The second purpose was to critically analyse the condition of the international reform evidence base. The analysis was completed in two parts. The first part concentrated on the quality of the evidence. It found that the methodological characteristics of the evidence associated with the five initiatives were more compelling than the magnitude of improvement. However, I argued that the small gains should not be trivialised. The second part of the analysis investigated the reasons for the condition of the evidence. Two practices were found to account for its high quality; investing adequately into summative evaluations and engaging expert programme evaluators to complete outcomes focused research. External accountability systems and role clarity were presented as priority constraints for those two practices. Those two constrains paved the way for expert programme evaluators to produce high quality evidence which evaluators then used to debate the value of the initiatives amongst themselves and with the developers.

In closing, the inquiry process used to identify the five international initiatives with the strongest evidence of effectiveness, in particular the criteria developed to categorise evidence as strong, promising or in need of further research, was extremely useful. Although it was time consuming and thought provoking in the making, the important task of selecting the initiatives became a relatively simple one of separating out those initiatives that had strong evidence from those that did not. It assured the objectivity that I was looking for in that it took away any opportunity to favour personal preferences. It also meant that links to evidence of effectiveness have been
made explicit from the outset and do not have to be assumed. Additionally, knowing that the five initiatives have high quality evidence of effectiveness provided confidence that the examination of their common development and implementation characteristics at the beginning of Part III would provide benchmarks for comparing what was happening in New Zealand’s most effective initiatives. The next chapter outlines the selection of the New Zealand schooling initiatives and explains the condition of the evidence attached to them.
Chapter Four: Effective Schooling Improvement in New Zealand
Deciding which initiatives should be used to design the New Zealand-based learning process model was not as straightforward as the decision-making process for the five international initiatives. In the international search, strong evidence was found for the five initiatives. That not only differentiated them from the rest but it also set a benchmark for finding equivalent evidence in the New Zealand context. The ideal in the New Zealand-based search was to find initiatives with strong evidence in order to make comparisons based on equivalent evidence. However, no New Zealand initiatives with strong evidence were found, which meant moving on with a less than ideal situation for comparisons.

The seven successful schooling improvement initiatives found to have the strongest evidence of effectiveness in the New Zealand literature are outlined in Table 6. The first five are cottage industry initiatives. To recap, cottage industry initiatives involve small localised groups of people designing initiatives to suit their local circumstances. The five cottage industry initiatives are; Achievement in Multicultural High Schools, Performance Enhancement in North Waikato (PEN), Students On A Rise in Paeroa (SOAR), Strengthening Education in Mangere and Otara (SEMO), Te Putahitanga Matauranga (TPM). In contrast, the other two initiatives in Table 6 are national professional development programmes attached to the Ministry’s curriculum division. They are the Numeracy Development Project & Te Kotahitanga.

Table 6. New Zealand Initiatives with the Strongest Evidence of Effectiveness.

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Evidence of gains in student achievement</th>
<th>Effectiveness rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A cottage industry initiative called <em>Achievement in Multicultural High Schools</em> (AIMHH) spread across the cities of Auckland,</td>
<td>A Ministry overview document of progress reports (Ministry of Education, 2005a) claims:</td>
<td>Needs further research</td>
</tr>
<tr>
<td></td>
<td>- a ‘sharply rising trend’ overall for the students’ external examination results in the National Certificate of Education</td>
<td>- the information in the</td>
</tr>
</tbody>
</table>
### Chapter Four: Effective Schooling Improvement in New Zealand

| Manukau and Wellington. - 9 Decile\(^{22}\) 1 secondary schools in NZ with high numbers of Maori and Pacific students. | Achievement (NCEA) from 2002 through to 2004; and  
- that norm referenced tests (asTTle scores) in years 9 and 10 show an improvement (subject areas not specified). | progress report has not been published  
- there was no calculation of statistical significance  
- there were no third-party or comparison studies. |
| --- | --- | --- |
| **Intervention**  
Started in 1996 to create strong college governance and management, strong school/community relationships and integrated social services. A coaching programme for teachers has recently been introduced. | A cottage industry initiative called *Performance Enhancement in North Waikato (PEN)* in the disadvantaged district of Huntly-Ngaruawahia in the Waikato region — 16 Decile 1-4 schools with a high proportion of Maori students | A progress report in a Ministry funding memo (Ministry of Education, 2005f).  
- The memo claims: “increases in stanine scores in reading comprehension tests [Supplementary Tests of Achievement in Reading — STAR (Elley, 2001)]... Of particular significance is that improvements have been made at both the critical levels (stanines 1-3) and at the higher levels (stanines 7-9). Some examples of the improvements are:  
  - STAR assessments across all year levels in 2004 show a reduction in the numbers in the critical areas of stanines 1-3 (average decrease of 14%) and an increase in the number in stanines 7-9 (average increase 8%);  
  - in the Year 3 and Year 4 cohorts, the reduction in the percentage scoring in... |
| **Intervention**  
Started in 2003 help teachers teach literacy more effectively. A literacy professional development programme provided teaching professionals with seminars and workshops to learn new intervention strategies as well as to... | Needs further research | --- |

---

\(^{22}\) A 'decile' is a socio-economic indicator to rate schools from 1 to 10 in order to allocate operational funding and staffing (Ministry of Education, 2006a). Decile 1 is the lowest rating and attracts the highest levels of funding and Decile 10 is the highest rating with the least amount of discretionary funding. From 1 January 2005 deciles have been calculated on five factors: household income - percentage of households with income in the lowest 20 per cent nationally, Occupation — percentage of employed parents in the lowest skilled occupational groups, household crowding — number of people in the household divided by the number of bedrooms, Educational qualifications — percentage of parents with no tertiary or school qualifications, Income support — percentage of parents who received a benefit in the previous year.
| Engage in analytical discussions about the impact of their interventions on students' reading ability. ICT support for literacy has been a secondary intervention. | Stanines 1-3 in 2004 was more than 23% and the increase in stanines 7-9 was by 20% in Year 3 and by 7% in Year 4;  
- in the Year 3 cohort those students scoring in stanines 7-9 in 2004 was increased by 18%" (p.5). | Needs further research  
- the progress reported was in only one study, which is yet to be published  
- there were no calculations of statistical significance  
- there were no third-party or comparison studies |
|---|---|---|
| A cottage industry initiative called Students On A Rise in Paeroa (SOAR) in the disadvantaged district of Paeroa in the Waikato region - 10 decile 1-3 schools with high numbers of Maori students | A Ministry overview document of progress reports (Ministry of Education, 2005d) makes claims of improvement in reading comprehension:  
- “Overall STAR results in 2004 for students in Years 4-9 show that: “there has been a positive shift in students moving out of the lower stanine groups and an increase in the percentage of students reading above their chronological age level (stanines 7-9). Overall the average stanine achieved has increased by 0.3 during the year. Students in year 4 have improved by almost 1 stanine over the year, while the average move in other year levels ranges from 0 to 0.3” (p.2).  
- Further breakdowns are provided which indicate Maori students are improving faster in comparison to all students: “The gap between Maori student results and overall results is closer – 25% difference in February 2003 and 8% difference in October 2004. Compared to the Total student results, there is a higher percentage of Maori students in the low and middle stanine groups, however there has been a positive shift in the percentage of Maori students out of the lower stanine group and into stanines 4-6” (p.3). | |
| A cottage industry initiative called Strengthening Education In Mangere and Otara (SEMO) in the two disadvantaged communities of Mangere and Otara in Manukau City – 37 decile 1-5 schools (mainly decile 1). | Nine studies have documented achievement gains in relation to two interventions: a reading intervention into classrooms of Year 1-3 students (Phillips et al., 2001; Timperley, 2002, 2004a; Timperley, Phillips et al., 2003); and another reading intervention into classrooms of Year 4-8 students (Lai et al., 2004; Lai et al., In Press; Lai et al., 2006; Lai et al., 2005; McNaughton et al., 2004). | Promising evidence  
- the student sample sizes in the first series of studies were small |
Predominantly Maori and Pacific student population.

**Intervention**

Started in 1997 with a literacy project designed to provide additional support to high performing schools and, in low performing schools, to address governance and management systems and leadership. More recently, several learning networks have designed and implementing reading interventions with Year 1-3 students and Year 4-8 students.

The first series reported on a sample of 343 students (108 five-year-olds, 135 five-and-a-half-year-olds and 100 six-year-olds) from 12 of 21 schools involved in the professional development programme

- the first pre-post study tracking Year 1 students in 12 of the 21 schools reported: “Significant gains, both statistically and educationally were achieved across a broad range of literacy measures” (Phillips et al., 2001, p.154). Eight measures were taken from the 6-year net Observation Surveys (Clay, 1993). Significant effect size gains from the eight measures ranged between 0.34 and 0.71;

- three subsequent comparison studies (Timperley, 2002, 2004a; Timperley, Phillips et al., 2003) used the text level and BURT scores from the schools’ Observation Surveys (Clay, 1993) to report that the significant gains that were made in the first study were sustained with successive cohorts of Year 1 students over the next three years: “the gains in the combined reading score were significant (t = 4.04, p<0.01) with an effect size of 0.28” (Timperley, Phillips et al., 2003). Two of the three studies were third party (Timperley, 2002, 2004a). The sole author of those two studies had not been involved in the original study.

The second series of studies involved;

- An initial study (Lai et al., 2004) collected baseline reading comprehension achievement information for 1975 students in Years 4 to 9 from seven decile 1 schools in the district of Mangere, “the means [in reading comprehension stanines] from Progress and Achievement Tests and STAR were 3.10 (SD = 1.38) and 3.10 (SD = 1.50) respectively. This indicates that the average student was just below the “average” band (stanines four to six) and was two stanines below the expected average of stanine five” (p.228)

- A second study (McNaughton et al., 2004) reported on phase one of the impact of an intervention designed and implemented to improve the reading comprehension of 1216...
students in year four to eight from six decile 1 schools in Mangere. The, “At each year level test comparisons [using STAR test] showed there was a significant gain in stanine, with effect size (using stanines) ranged from .42 to 1.37. There was an overall gain in achievement in stanines across year levels of 0.52 of a stanine. This meant that relative to the baseline pattern which included that students gained about one year, for a year at school students gained about 18 months (p.190).

- A third study (Lai, 2005) reported on the second phase of the intervention to improve the reading comprehension. It involved the 586 year four to eight students from six decile 1 schools in Mangere who were also involved in the phase one study. The study claimed, “there was a statistically significant increase in achievement [using the STAR test] between Time 1 (Feb 03) and Time 4 (Nov 04) in every cohort, with an overall gain of 0.8 of a stanine. Given that stanine scores are age-adjusted, the overall gain of 0.8 stanine indicates that the intervention has advanced student achievement across all cohorts by approximately nine months in addition to the expected national advancements made over two years” (p17).

- A fourth study (Lai et al., 2006) reported on the results of replicating the reading comprehension intervention conducted in the Mangere schools with 1161 year four to eight students in six schools in the district of Otara. The study reports, “Across year levels, there was an overall average gain in achievement of 0.81 of a stanine [using STAR test results] with an overall effect size gain computed from standing scores of 0.50. Given the stanine scores are age adjusted, the overall gain of 0.81 indicates that the intervention was associated with gains in student achievement across the total group by approximately nine months in addition to the expected national advancements made in a year.

- A fifth study (Lai et al., In Press) reported on the 3-year reading comprehension intervention in Mangere in 7 schools. Cohorts tracked over three years (n=238) showed statistically significant overall acceleration in achievement from Time 1 to Time 6 of 0.97 stanine (STAR). This
The Numeracy Development Project — a professional development programme within the National Literacy and Numeracy strategy since 2000. By the end of 2005, the mainstream and Māori seven studies have documented achievement trends for various parts of the project (Higgins, 2001, 2002; Higgins et al., 2004; Thomas & Tagg, 2005a, 2005b; Trinick & Stephenson, 2005; Young-Loveridge, 2005). The assessment tool was a diagnostic interview between the teacher and student to provide teachers with a detailed knowledge of student needs further research.

Needs further research.

The progress report is not part of a study to be published.

There was no calculation of statistical significance of claims made.

There were no third party or comparison studies.
medium components would have operated in approximately 1,800 schools across the decile range.

**Intervention**

Started with 12 trial schools in 2001. It is a national professional development programme for teachers. A number framework and a diagnostic survey provide a guide for lesson planning. Teachers are encouraged to use three types of lessons (knowledge lessons, strategy lessons and a combination of knowledge/strategy lessons) and to complete follow-up diagnostic surveys to check progress.

- The latest evaluation report stated that: “All students seemed to benefit from participation in the project, but some groups made greater gains than others. Asian students were the highest performers... Pakeha/European students did better than Maori students, who in turn outperformed Pasifika students. However, there was evidence to suggest that the gaps between groups may be getting smaller” (Young-Loveridge, 2005, in Higgins et al, 2005).

- Results in addition and subtraction across all year levels appeared slightly better than for the other operational domains (multiplication/division & proportion/ratio). In an analysis of 2004 results, one researcher claimed: effect sizes were large [in year groups 1 & 2] (0.83), but diminished to medium by about year 4 (0.58 to 0.34), and small (0.20) by year 8. It is likely that ceiling effects helped to reduce the magnitude of effect sizes in the senior primary and intermediate years” (Young-Loveridge, 2005, in Higgins et al, 2005, p.9)

- Researchers in another study of students in 31 schools involved in the numeracy project found that: “The schools who reported extensive use of numeracy achievement data appeared to raise the achievement of their students more than schools with a lower reported use of achievement information” (Thomas & Tagg, 2005a, in Higgins et al, 2005, p.21)

**Te Kotahitanga – a national professional development initiative operating in 12 secondary schools across New Zealand.**

**Intervention**

Started in 2001 to improve teachers’ relationships with Maori students in mainstream secondary schools. The programme

- One study (Bishop et al., 2003) reported “some noticeable improvements in achievement of Maori students in the classes taught by some of the teachers in this project” (p.153). “E.g. Maori students in [a mainstream science class] 10NH achieved positive academic results only in Science, which was the subject taught by one of the participating target teachers” (p.159).

**Needs further research**

- There was only one published study

- There was no indication of statistical significance of the positive gains
4.1. Selection of New Zealand Initiatives for Further Analysis

Selecting initiatives in New Zealand for further analysis was no straightforward matter. It took quite a complex decision making process to select the SEMO project and the Numeracy Development Project. An immediate decision was made to select the SEMO-related learning networks for further analysis because the evidence was rated ‘promising’. That rating set it apart from the others as the evidence bases attached to the other six needed further research.

Evidence attached to the SEMO project that got it the promising rating was found in nine empirical studies, which were conducted in two series. One series of four studies tracked the impact of a reading comprehension intervention for successive cohorts of Year 1 student in 12 schools in the first study (Phillips et al., 2001) then in seven of the original 12 schools over three consecutive years (Timperley, 2002, 2004a; Timperley, Phillips et al., 2003). The primary intervention strategy was professional development to improve the teachers’ instructional practices in reading and reading comprehension (Phillips et al., 2001). Small samples of approximately 100 Year 1 students were involved in the first study and then smaller samples again in the three subsequent comparison studies. An important part of tracking improvement was that the researchers (who were also the developers so not independent) negotiated with the participating schools to use generic standardised tests to measure the students’ reading comprehension before and after the intervention. The tests were eight measures taken from the Six-Year Net Observation Surveys23 (Clay, 1993) and the BURT word reading test (Gilmore, Croft, & Reid, 1981). The researcher-developers’ measures taken at the beginning and end of the intervention in 12 schools

---

23 The eight measures were; concepts about print (CAP), letter identification (LID), word recognition (WORD), writing vocabulary (WRVOC), Hearing and recording sounds (H & RS); Story Retelling (RETELL) and the Peabody Picture Vocabulary Test (PPVT)
found effect size gains ranging from 0.34 to 0.71 for the eight measures (Phillips et al., 2001).

A third-party researcher conducted three comparison studies in the three years after the original study in seven schools (Timperley, 2002, 2004a; Timperley, Phillips et al., 2003) of which the 2002 and 2004 studies had third-party status. The three comparison studies investigated whether new cohorts of Year 1 students were performing as well or better than the cohort involved in the original study. A key finding from the three comparison studies was that the gains were not only sustained but continued to improve for all the schools except one, which dipped in the fourth year (Timperley, 2004a). The dip did not lower the overall significant improvement below the initial baseline achievement gains recorded in the first study.

It was the addition of the second series of five studies that lifted the rating to promising. Four of the studies investigated the impact of a classroom intervention to improve reading comprehension among Year 4-8 students in six Mangere schools over three years (Lai, 2005; Lai et al., 2004; Lai et al., In Press; McNaughton et al., 2004). The fifth study reported a replication of the reading comprehension intervention in six schools in the neighbouring district of Otara (Lai et al., 2006). In this case it aimed at improving the teachers’ skills in analysing and using achievement information in order to alter classroom teaching practices in reading and reading comprehension. The researcher-developers used a standardised test appropriate for the student age group, i.e. Supplementary Tests of Achievement in Reading (STAR) (Elley, 2001). Student sample sizes in this series of studies were much larger than those in the first series. Baseline data was collected for 1975 Year four to eight students from six schools in Mangere. Achievement data in the three comparison studies in Mangere came from 1216, 586, and 238 students respectively. The reason why the sample sizes reduced in each successive study was because the researchers only reported on students who were involved in the original study which collected baseline data. In the replication study in Otara, the sample size was 1161 students.

To act consistently with the decision-making process used to select the international initiatives, I should have selected the SEMO project and left the New Zealand-based search at that. However, that choice would have created two dilemmas. One dilemma was that the use of one cottage industry initiative such as the SEMO project for the investigations in Part III did not reflect the mix of schooling improvement initiatives operating in New Zealand. The contents in Table 6 verify a
mix of cottage industry and national initiatives. The second dilemma was to avoid a potential personal bias towards cottage industry initiatives. My considerable personal investment into the SEMO project from 1997 through to 2002 and since then my role in helping develop other cottage industry initiatives around the country meant that I was prone to favour that type of school improvement.

With those dilemmas in mind, I had a difficult decision to make. I could proceed consistently with my criterion-based selection principles and select the SEMO project as the only New Zealand initiative worthy of further analysis. Alternatively, I could compromise the selection principles and select one more, a few or all of the six remaining initiatives to ensure that the investigations in Part III were not dominated by findings from the cottage industry approach. I decided it was better to present a more balanced picture than adhere rigidly to the criteria for selection for two reasons. One reason for doing so was because the overall aim of this thesis is to inform New Zealand’s schooling improvement movement. It did not make sense to only inform localised developments when I had identified some national initiatives in Table 6 that were also doing important schooling improvement work.

A second reason for compromising my selection principles was because the quality of SEMO’s evidence only just made the ‘promising’ rating in my view. Small to moderate sample sizes were my main cause for concern. Nash (2003) noted the questionable nature of the samples of students in the first series of studies. He was concerned about the high attrition rate in between the times that baseline and post-intervention achievement information was collected. Although the second series of studies involved much larger sample sizes, high attrition rates were problematic in that series of studies as well (1975, 1216, 586, & 238). In this second series of studies, the authors acknowledged the problem, “Only 238 students could be tracked over three years” (Lai et al., In Press).

Other researchers also critiqued the condition of the evidence attached to the two series of interventions (Harker, 2003; Tunmer, Prochnow, & Chapman, 2003). Harker (2003) as well as Nash (2003) critiqued the first two studies conducted (Phillips et al., 2001; Timperley, Phillips et al., 2003). Among a number of concerns, they were both critical that the evidence base did not justify some of the claims made in the studies. One particular claim that appeared to cause them considerable concern, and which is relevant to this section, was made in the original report, “by the end of the first year of school their [pupils in Decile 1 schools] achievement can be like any
other child in New Zealand” (Phillips et al., 2001, p. 191). Evidence supporting this claim were aggregated test results (6 Year Survey) which showed that the intervention students overall had entered the ‘average zone’. The critics believed the researchers’ interpretation of ‘average’ was overly generous. Harker (2003) elaborated on this concern by focusing on the percentile cut-off point between stanine 3 and 4.

In the literature, Stanines 4, 5 and 6 are regarded as “average”, and they contain 54% of the distribution of scores within a population. Having a distribution of schools that centre on the 23rd percentile point (the cut-off point between Stanines 3 and 4) is hardly the same thing as being “like any other child in New Zealand” or even “close to the expected levels”. To support such a claim, the scores of the intervention pupils would have to centre around the 50%. (Harker, 2003, p.246)

These concerns coupled with the relatively small sample sizes served as a caution not to single out the SEMO project, but to recognise it as the initiative with the most advanced evidence base in New Zealand. Hence, my justification for a more balanced sample of New Zealand initiatives for analysis in Part III.

My two dilemmas about assuring a balanced picture and avoiding a potential bias towards the cottage industry approach caused me to prefer including one of the national professional development strategies. As it happened, the evidence attached to the Numeracy Development Project was more developed than that attached to Te Kotahitanga, which was the only other nationally-driven professional development programme considered for selection. Whereas Te Kotahitanga only had one research report which included case studies of improved student achievement (Bishop et al., 2003), the evidence base for Numeracy Development Project included seven studies which documented achievement trends for various parts of the project or made reference to the achievement trends (Higgins, 2001, 2002; Higgins et al., 2004; Thomas & Tagg, 2005a, 2005b; Trinick & Stephenson, 2005; Young-Loveridge, 2005). The assessment tool used to measure success in the Numeracy Development Project was a diagnostic interview between the teacher and student (Ministry of Education, 2004b). It served a dual purpose; to provide teachers with a detailed knowledge of student thinking in order to design relevant lessons, and to provide project leaders with collated diagnostic findings to inform the design of the next stages of the project (Higgins & Parsons, 2005).
Results in addition and subtraction across all year levels appeared slightly better than for the other operational domains (multiplication/division & proportion/ratio). In an analysis of 2004 results, one researcher claimed that, “effect sizes were large [in year groups 1 & 2] (0.83), but diminished to medium by about year 4 (0.58 to 0.34), and small (0.20) by year 8. It is likely that ceiling effects helped to reduce the magnitude of effect sizes in the senior primary and intermediate years” (Young-Loveridge, 2005, p.9). Researchers in another study of students in 31 schools involved in the numeracy project found some variations in the results across schools (Thomas & Tagg, 2005a, in Higgins et al., 2005, p.21). Their analysis found that, “The schools who reported extensive use of numeracy achievement data appeared to raise the achievement of their students more than schools with a lower reported use of achievement information” (p.21).

A problem I have with the information in the above paragraph, which also helps explain why I did not rate the evidence as promising, is that it is based on teacher perceptions of students’ thinking recorded from diagnostic interviews. I do not consider that a diagnostic tool of that nature can produce robust results about students’ academic achievements as standardised tests can. It is extremely useful formative information but tenuous as the only base line and follow-up data used to measure the magnitude of improvement. Feedback from one policy developer indicated that a triangulation exercise has been carried out to verify the positive results in a district called Manurewa, which is close to Mangere and Otara (National senior policy advisor for numeracy, New Zealand Policy Developers’ Feedback, 2005). A norm-referenced test (asTTle) was used alongside the diagnostic interview. Although the results of the exercise have not yet been published, a recent report indicated that the exercise not only validated the positive diagnostic findings but also found that the schools in Manurewa were doing better than similar schools nationally (Young-Loveridge, 2005). I considered these findings particularly interesting and decided to explore the developments further in Part III (6.3.3.).

To summarise, the SEMO project and the Numeracy Development Project were selected for further analysis in Part III. Selecting those two initiatives was not easy. From a purist research perspective of adhering to outcomes-focused methodologies, it would have been ideal to have at least two initiatives with strong evidence of effectiveness. However, that was not to be. Evidence attached to the SEMO project
was promising and that in relation to the Numeracy Development Project needed further research. In the interest of informing New Zealand’s schooling improvement movement, I had to work with the most successful local and national initiatives available and those two initiatives were the ones identified.

4.2. The Condition of the Evidence in New Zealand

The summary of the last section alludes to my view that the overall quality of the evidence attached to the initiatives in New Zealand is low. This section explores that view in some detail with the intention of explaining the reasons for the low quality.

4.2.1. The quality of the evidence

The evidence of effectiveness attached to all seven initiatives in New Zealand considered in the selection process needs considerable development before it can be considered high quality. That claim includes the most advanced evidence bases. The previous section outlined developmental issues about insufficient sample sizes and overly-generous interpretations of success that were attached to the SEMO evidence base (Harker, 2003; Nash, 2003). In relation to the Numeracy Development Project, more triangulation of the results from the diagnostic interviews and the introduction of third-party research projects would help strengthen the evidence base. Evidence attached to the other five initiatives reported in Table 6 (Ministry of Education, 2005a, 2005c, 2005d, 20050 had two noticeable shortfalls. The first shortfall was a lack of statistical analysis of achievement gains. Not all the initiatives used measurement tools that could generate a calculation of statistical significance. Those that did use such tools did not develop the results much beyond organising them into graphs and tables. The second shortfall was a lack of third-party research to verify the results claimed by the researchers and developers. The shortfalls did not indicate a lack of suitable achievement information or an unwillingness to further develop the evidence. To the contrary, with regard to the first point, there appeared to be ample data to make judgements of effectiveness. With regard to the second point, at the time of this investigation the leaders of the initiatives indicated that they were in the process of further developing their evidence.
4.2.2. Reasons for the underdeveloped condition of the evidence

Why is it that 15 years after developed countries moved into an outcomes-focused era (Fuhrman, 2002), New Zealand's school reform evidence base remains underdeveloped compared to the evidence found in England and the United States? The constraints analysis in Table 7 and the explanations in this section help to answer this question.

Table 7. Constraints Analysis for the Underdeveloped Condition of Evidence Attached to the Two New Zealand Initiatives.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Practices</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>• National policy silent on evaluating schooling improvement initiatives operating across schools</td>
<td>• Diverse methodologies to evaluate effectiveness</td>
<td>• An under developed school improvement evidence base</td>
</tr>
<tr>
<td>• Fierce professional opposition to centrally directed standardised testing</td>
<td>• Heavy investment into development and little in programme evaluation</td>
<td></td>
</tr>
<tr>
<td>• A New Zealand cultural norm of getting on with the job known as the No.8 wire approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Limited systems for collating and using achievement information across schools</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.2.1. Diverse methodologies to evaluate effectiveness

The first practice that contributed to an underdeveloped evidence base is the diversity of methodologies used by leaders of the school improvement initiatives to evaluate effectiveness. Progress reports to the Ministry indicated that none of the reform groups followed the same pattern in establishing their claims of success (Higgins et al., 2005; Ministry of Education, 2005a, 2005c, 2005d, 2005e, 2005f).
One of the seven initiatives used national tests (AIMHI) and of the other six, four used norm-referenced tests of their choice (SEMO, Huntley-Ngaruawahia, Paeroa & Te Putahitanga Matauranga), one developed a set of diagnostic surveys (the Numeracy Development Project) and one used a case study approach (Te Kotahitanga). The seven initiatives then went on to develop their claims using different pathways. For instance, the leaders of the SEMO-related initiatives chose to buy into replication studies and third-party comparison studies whereas the leaders of the Numeracy Development Project continued using their diagnostic survey. The leaders of each initiative are effectively creating their own unique evidence base. Consequently, generating a common calculation of improvement, such as an overall effect size gain, is reliant on the initiative leaders agreeing to and using some common evaluation practices. To date agreement on and use of common evaluation practices has occurred within those initiatives with the strongest evidence of effectiveness but it has not occurred across them.

Feedback from senior policy developers indicated that there are no policy constraints stopping that from happening (National managers with oversight for national professional development programmes and cottage industry initiatives, New Zealand Policy Developers' Feedback, 2005). So what was constraining robust programme evaluation across schooling improvement initiatives as an alternative to an acceptance of diverse methodologies? I suggest a main contributor to the inaction was a policy silence in that area of evaluation. Other forms of evaluation have been prioritised for national policy development over the past decade. Government could have imposed national testing, notwithstanding the self-management law, but it had a policy platform not to do so (National manager with oversight for cottage industry initiatives, New Zealand Policy Developers' Feedback, 2005). One form of evaluation of particular policy importance has been in-school formative assessment practices (Ministry of Education, 2001). The national policy preference was for school boards and practitioners to focus their evaluative attention within their schools rather than across them. Teachers, in particular, were encouraged by the national assessment policy to keep their assessment focus within their classrooms and develop formative assessment relationships with their students. Meanwhile, assessing what was happening across schools was determined by the various national and international surveys mentioned in Chapter 2 (2.3., p.32).
The thinking behind those policy preferences appeared to be that if schools took the lead, then they were more likely to want to solve the problems than if they were told to do so (Sinclair, 1999a, 1999b, 2001). That thinking is useful in lifting the overall capacity of schools to respond to achievement challenges as they arise. However, in relation to schooling improvement in disadvantaged districts, it relied on the school leaders accepting that they had an underachievement problem to solve. This is problematic because it is taking considerable time for many self-managing schools serving disadvantaged communities to accept the seriousness of the underachievement problem that exists. A comment from a principal quoted in a review of the monitoring and support division captured the feelings of many self-managing school leaders facing up to the problem, “You feel like you’re trying so hard all the time and it’s very hard to say it’s not working... so it’s a barrier even asking for help and identifying the real issues” (McCauley & Roddick, 2001, p.70). Changing that sort of attitude is particularly important given that their involvement in schooling improvement initiatives is mainly voluntary. If school leaders do not recognise that there is an underachievement problem then they are unlikely to commit fully to initiatives aimed at solving the problem.

A second contributor to the lack of programme evaluations across schools is that existing national assessment policies did not engender a sense of urgency to find out what was and was not working across schools. An analysis of what happened in the recent and distant past helps to explain this second contribution. In terms of the distant past, current attitudes within schools towards student achievement still appear to be influenced by assessment practices that were widespread prior to the 1989 Tomorrow’s Schools reforms. What was important in assessment in schools prior to 1989 was to fit students into a bell-shaped curve as an accountability task to principals and inspectors. The bell-shaped curve has most cases in the middle of the curve and, at either side, the curve drops away, first slowly then more rapidly, with a relatively long tail at either end (Thorndike & Hagen, 1977). Having all students appropriately placed on the curve was a sign of effective assessment24 (Department of Education,

24 A key schooling practice to place all students appropriately on the bell-shaped curve was to assess students’ achievements on two occasions each year using a five-point scale: 1 (Outstanding), 2 (Above Average), 3 (Average), 4 (Below Average) and 5 (Extremely limited). The Department’s notes about the ratings encouraged teachers to think about the right balance in keeping with the bell-shaped curve: “A teacher would expect that out of a representative sample of one hundred such pupils: five would receive the top rating of 1; twenty would receive a rating of 2; fifty, or about half, would receive a rating of 3; twenty would receive a rating of 4; and five would receive the bottom rating of 5”
1987). The exercise was completed at the end of each school year. Theoretically, that exercise, completed properly throughout the country, would produce a normal distribution of the entire student population. The purpose of the exercise was to classify students. It was not to improve teaching. There was no systematic analysis of the distribution. Instead, each school stored their achievement information safely in the unlikely event that the Minister of Education chose to authorise its destruction.

The way in which the annual assessment ratings of students were completed in some disadvantaged communities masked the underachievement problem. Researchers who analysed the assessment procedures in the districts of Mangere and Otara, for instance, found that the schools in those two districts were creating an impression that student achievement was higher than it really was (Ramsay, Sneddon, Grenfell, & Ford, 1981). They tended to compare the academic performance of students against the school population rather than the national population. Their ratings inflated the performance of the students. At the end of each year, the principals and inspectors generally accepted the inflated ratings rather than requiring a more accurate representation of how well students were doing. Many of the ratings would have been shifted down the scale had the teachers followed the Department’s guidelines correctly \(^{25}\) and had the principals and inspectors verified the ratings in line with national norms. Had that happened, it is possible that the Department and the schools as well as the families and the students might have become more concerned about the underachievement problem in the two communities. That prediction, however, may be overstating things given that students living with disadvantage were considered desirable for the manual labour market and that market was vibrant prior to the Tomorrow’s Schools legislation (Butterworth & Butterworth, 1998).

In the more recent past, the new assessment processes established after the Tomorrow’s Schools legislation still did not arouse a sense of urgency. Ratings against the bell-shaped curve were replaced by ratings of achievement against levels

---

\(^{25}\) The guidelines stated that the spread of assessments in any given class depended on the mix of students (Department of Education, 1987). For mixed ability classes which reflected the national student population, the spread was expected to be similar to the percentages outlined in the guidelines. They also explained that some mixed ability classes would vary considerably from those percentages and that high-ability and low ability classes would tend to be bunched towards the upper or lower end of the curve.
specified in the national curriculum (Ministry of Education, 2004d). The levels provide generous timeframes for students to progress through the framework. Whereas the ratings were checked annually by inspectors prior to the 1989 reforms, the new system saw the levels subject to accountability reviews approximately every three years by the Education Review Office. The schools’ attempts to assess students’ levels and the Review Office’s accountability checks of the accuracy of those attempts have also not been used as primary evidence to activate interventions into struggling schools under the Ministry’s intervention framework (M. O’Carroll, personal communication, July 4, 2006).

These national assessment policy changes could have a number of implications. They could mean that schools are continuing to inflate their ratings and that those ratings are being accepted in much the same way as occurred in the past. A scenario written by the research team tracking the SEMO project about a typical disadvantaged girl growing up in Mangere and Otara implied this to be the case (Robinson & Timperley, 2004). The scenario entitled ‘Lita’s Story’ is about a fictitious character called Lita whose experience of schooling was considered by the researchers to be common to many students in the disadvantaged districts. It explains how she was able to progress through her primary and middle school years thinking she was doing really well at school and her parents were told likewise through regular school reports. When Lita went to high school she aspired to go to university and become a manager of a credible company. However, to her and her parent’s dismay, she got poor reports and failed the external examinations. Rather than laying the blame on the performance of the high school, the researchers questioned the feedback that students such as Lita and those students’ families were receiving through her earlier years at school: “In all but two [of the 11] schools, parents did not have access to accurate information about the achievement of their children” (Robinson & Timperley, 2004, p. 67). The Lita’s story publication includes quantitative and qualitative data to show that the inflationary practices Ramsey and colleagues (1981) reported on rather than the fictitious story was the stimulus for formal school reports to students and parents.

The important point here is that although the national assessment system was altered in the early 1990’s and additional related policies added to the mix, such as the Revised National Education Guidelines (Ministry of Education, 1999) and the Education Standards Act (New Zealand Government, 2001c), many successive cohorts of disadvantaged students and their families have meanwhile continued to
have inflated opinions of their academic ability throughout much of their schooling careers. That would suggest that the evaluative levers at the systems level have fallen short of helping school leaders and practitioners working in disadvantaged communities to recognise the overall underachievement problem and to take some responsibility for ensuring the students and their families are aware of it.

In summary, a policy preference in evaluation has been to play a long game and give schools control over summative evaluation so that they respond to achievement challenges as they arise. The problem with the long game is that schools have been slow to recognise the extent of the underachievement that needs to be addressed. Past and present systems for rating student achievement appear to have contributed to the slow recognition. This situation is particularly problematic in terms of evaluating schooling improvement initiatives. Had school leaders in Mangere and Otara recognised the extent of the underachievement problem in their schools earlier they may have demanded exposure to the sorts of robust evaluations associated with the international initiatives. Instead, they have helped create unique programme evaluations for each initiative. Another possibility put forward by a local official working with the Mangere learning network which is in keeping with the No.8 Wire approach is that the school leaders would have pressed for more action to solve the problem, but not necessarily evaluation of that action (Warren, 2006). A consequence of the long game at this stage is that the quality of the evidence is low. Should the long game succeed and the school leaders eventually recognise the full extent of the achievement problem, they are likely to demand higher quality evaluations to check on their progress in addressing it.

4.2.2.2. Investing in development over research.

The second practice that has contributed to the low quality of reform evidence in New Zealand is the heavy investment in reform development tasks and the relatively low priority given to evaluating their success. There is unquestionably a great deal of development going on. The search for evidence of effectiveness found around twenty cottage industry initiatives currently operating and numerous national levers purporting to help solve the problem of underachievement among students from disadvantaged communities. However, on the research front, which is the focus of this section in terms of programme evaluation in schooling improvement, things are
not as active. There are two parts to this phenomenon, a bigger picture part and a part specific to schooling improvement initiatives.

In terms of the bigger picture, overall investments into research do not appear to have produced a sufficiently capable research community to conduct high quality programme evaluations of schooling improvement initiatives. One senior policy developer (Alton-Lee, 2005) referred me to findings in a recent OECD report (OECD, 2003a) that New Zealand is well behind other developed countries in linking research to development, "New Zealand invests far less in research and development of any kind than other developed countries, and has far lower research and development personnel per million population than Australia or Western European countries." (p.89). This point was reinforced recently by several New Zealand researchers who formed a peer review panel to provide feedback about the performance of New Zealand's research community in the first education Performance-Based Research Fund (PBRF) evaluation (Alcorn et al., 2004). They stated, "The overall research strength of education academics in New Zealand, however, would have to be regarded as weak. Around three-quarters of eligible staff in this discipline did not attain a research-active rating" (p.296).

In terms of specific research into schooling improvement, there have been some investments into research but there is far more emphasis on informing the development of the initiatives than assessing their effectiveness. In the SEMO project, for instance, approximately 15.6 percent of additional government funding spent on the project from 1998 through to 2002 was spent on research and evaluation26 (Annan et al., 2004). Only about a third of that investment was put into outcome-focused research (Phillips et al., 2001; Timperley, 2002, 2004a; Timperley, Phillips et al., 2003). The rest went into Education Review Office accountability reports and formative research reports so that school leaders supported by Ministry officials could prioritise development needs. Therefore, about 5 percent of the additional funding was spent on outcomes-focused evaluation.

---

26 Investments into SEMO included $7,049,738.15 (NZ) for development and $1,103,839.2 for research and Evaluation from 1998 through to 2002 (Annan et al., 2004). That translates to 84.4 percent invested into development and 15.6 percent invested into research and evaluation. Investments into the Numeracy Development Project have been approximately $9.73 million per year from 2002 to the current financial year (M. Hyland, personal communication, July 6, 2006). Investment into development is 96.1 percent and research and evaluation is apportioned approximately 3.9 percent.
I propose that two constraints have influenced the lack of investment into outcomes-focused programme evaluations of school improvement initiatives in New Zealand. One constraint relates to the New Zealand cultural norm of just getting on with the job which I referred to at the beginning of Part I as the “No.8 wire” approach. The quality of the mend is not important. It is a matter of getting on with things. What I am suggesting is that reform groups in New Zealand prefer to get on with things rather than subject themselves to robust checks of effectiveness. One senior policy developers’ feedback concurred with this suggestion in acknowledging a link between New Zealand’s schooling improvement developments and the No.8 wire cultural norm. She noted that the approach has advantages and disadvantages, “On the positive side it does get things done and it is low cost. However, it is usually neither sophisticated nor replicable in different circumstances” (Eppel in New Zealand Policy Developers’ Feedback, 2005, p.6).

Through a broader analysis of historical literature, it became apparent that the approach has been a strong influence on the New Zealand education system for a considerable length of time. There are parallels between its use in developing evidence of effectiveness for schooling improvement initiatives and its use in developing the education system in the past. Prior to the 1989 reforms, for instance, the No.8 wire approach appears to have been used to sort most things out for over a century. One researcher points out that New Zealand’s first Education Act in 1877 had serious inadequacies (Butterworth, 1993) which were addressed by a series of “fixes” rather than legislative reform. The 1877 Education Act (New Zealand Government, 1877) set in place free, secular and compulsory schooling policy for students up to the age of 15 years. However, it did not make provision for the vast array of educational issues that were to come up in the 112 years that unfolded before that landmark legislation was comprehensively reviewed in 1989. In the absence of regular legislative reviews and revisions, the Department of Education just got on with the job of making it work. This resulted in a mountain of regulations that tried to deal with every eventuality. The regulations as well as the vast bureaucracy that grew to administer them were cited in the policy agenda setting process in the lead up to the 1989 reforms as a significant reason for the demise of the old system (Picot Taskforce, 1988).

This historical insight reinforces the senior policy developer’s quote about the No.8 wire development approach lacking sophistication. It may be a useful approach to
initiate developments when there is little known about what to do or how to do it. It is also useful for maintaining developments that do not require a great deal of sophistication. However, the approach is not useful in dealing with complex educational challenges such as evaluating schooling improvement initiatives. The two initiatives considered for further analysis in this chapter are examples of New Zealand educators moving away from the No. 8 wire approach in that regard. The leaders of those initiatives have chosen to engage researchers with expertise in evaluating the effectiveness of their initiatives. That expertise helped develop the outcomes-focused evidence, although the small scale and relatively unsophisticated methods compared to international methods tend to suggest that the No 8 wire approach is simply being employed in a more sophisticated way.

A second constraint influencing limited investment into outcomes-focused research is the difficulty of producing that sort of research without a systemic approach to collating and using achievement information. That point was made clear by the chief reviewer of the Education Review Office in an area-wide accountability report of a disadvantaged district early in the development of New Zealand’s school improvement movement (Education Review Office, 1997). The absence of any such system is attributed partly to the policy decision not to introduce national testing below the last three years of secondary schooling. A benefit of that decision is the absence in New Zealand of the traps that come with large scale achievement information systems attached to national testing programmes in England and state-wide and city-wide testing in the United States. Common traps avoided in New Zealand are teaching to the test within schools and claims made about schools’ performance based on the results of a few annual pencil and paper tests (Crooks, 2003).

There is also considerable professional suspicion that any attempts to build a system to manage achievement information is a cover for introducing national testing through the back door (Crooks, 2003; Flockton, 2003, 2005). Their arguments appear to be based on an assumption that any assessment other than national student sampling coupled with school and teacher-based formative assessments comprises a threat of dangerous external accountabilities. Such suspicion leads to delay, if not hostility, towards the development of new achievement management systems. There is a need for some restorative work to deal with the suspicions because there is little, if any evidence to substantiate the suspicion.
Chapter Four: *Effective Schooling Improvement in New Zealand*

Despite the avoidance of national testing and general suspicions, senior policy developers once again point out that there is no law or national policy stopping leaders of New Zealand’s schooling improvement initiatives from developing an achievement information system to track the success of their initiatives (National managers with oversight for national professional development programmes and cottage industry initiatives, New Zealand Policy Developers’ Feedback, 2005). Some systems have been developed by the leaders of some schooling improvement initiatives and the one attached to the learning network in Mangere is showing signs of becoming particularly sophisticated. That particular system is worthy of further explanation here because it may be the beginnings of the sort of system that the chief review officer of ERO was envisaging. Table 10 presents information relevant to its development as a three-phased scenario which unfolded over the past decade. The three phases were a ‘realisation’ phase (1998-1999), a ‘mastery learning’ phase (2000-2003) and a ‘transforming into practice’ phase (2004-present day). I developed these three phases after analysing documentation related to the achievement management system in Mangere several times.

In brief, the principals and middle managers leading the learning network in Mangere realised that they needed a robust system to manage their achievement information in order to conduct outcomes-focused evaluations of their efforts (Phase 1). They then embarked on a rigorous learning programme to grow their knowledge in programme evaluation (Phase 2) and built a system that they could use with the help of their preferred expert researcher-developers (Phase 3).

Table 8. Scenario: Developing an Achievement Management System in Mangere.

The *realisation phase* was about school leaders realising that there were serious achievement problems. They came around to those beliefs as a result of developing literacy-based projects as part of the SEMO project. The projects utilised additional government funding provided to solve the multitude of school performance problems identified by the Education Review Office (Education Review Office, 1996). Researchers monitoring the literacy projects reported at the end of the first year that many schools were adding-on to an already cluttered approach to literacy. They said most schools focused on “Coming up with a project, rather than on critical analysis of current practice... In short, many local people saw the government’s role as providing extra resources to those who knew what was best for the area and its students” (Timperley, Robinson, & Bullard, 1999, p.96). The Ministry’s local project team developed an outcomes-focused criterion for funding approval for the second year of the projects (Annan, 1999), which helped many schools change tack. By the end of the second year, most school leaders accepted that they had been using a data-free approach to try and...
improve things (Robinson, Timperley, & Bullard, 2000). The previous culture of complacency amongst the school leaders started changing to a more analytical culture of inquiry centred on finding out which instructional practices were working and which were not (Timperley, 2002).

The *mastery learning phase* started with a group of middle managers and principals from 11 schools forming a management team deciding to co-ordinate professional development in inquiry-based learning. Evaluating programmes was an important part of the programme, but it was not separated out from planning and implementation. They decided to contract a facilitator to help them but during the contracting process they realised a facilitator was unlikely to assist them to get beyond procedural tasks of planning, implementing and evaluating their literacy programmes. The chairperson of the management team reflected on the appointment process at the International Congress of School Effectiveness and Improvement conference in Sydney in 2002. She said:

“We felt as if we were just doing the same thing again, i.e. reinventing bureaucratic processes that kept us busy on procedures rather than interrogating teaching, management and governance practices through analysing student achievement information. Instead of a facilitator, I realised we needed a teacher, an analyst, a problem solver, a research literate individual…. We needed someone to challenge our assumptions, develop our skills in using achievement information, expand our thinking and enable us to become evidence-based decision makers” (Mose & Annan, 2002, p.4,5 & 6).

They found a postgraduate university lecturer who fitted their criteria. She helped them further shape the role she was to fulfil so that analysing problems in such a way that existing practices are critiqued (Robinson & Lai, 2006) was the first step in any planning process. Consequently, she was called a ‘problem analyst’. She proposed that the middle managers and principals undergo mastery learning and tests of key evaluation terms and concepts (Ministry of Education, 2003a). They accepted that challenge as well as her suggestion to develop protocols for handling achievement information across schools. That latter exercise proved to be particularly challenging in the self-managing environment. For instance, community meetings highlighted concern among Maori (indigenous) parents that their children’s achievement results be handled respectfully. Subsequently, it took several additional months work to clarify rights and responsibilities among school boards, practitioners, parents and students in relation to achievement information. The mastery learning and protocols helped to set some standards and to reduce any residual complacency among the school leaders about the extent of their knowledge and skills in programme evaluation.

The third phase of *transforming into practice* involved the problem analyst coaching the principals and middle managers in their schools how to evaluate the effectiveness of their practices. Coaching in this instance meant the principals and middle managers conducting case studies to prove their competence in evaluating what they were doing (Lai, 2003). They had to identify a data set then work through an assessment process\(^{27}\) that would improve what was happening in a classroom. In addition to the coaching, the problem analyst also helped the management team start working out a way of managing achievement information more efficiently. They had been wasting a lot of valuable analysis time through the use of painstakingly slow manual collations of information which tended to be error ridden (Ministry of Education, 2003d). They wanted an automated system capable of sophisticated

---

\(^{27}\) The assessment process involved nine steps: (i) analyse the data to identify above or below performance against agreed criteria (ii) explain the meaning of the data (outliers & patterns) (iii) analyse the strengths and weaknesses of the data (iv) identify results that need to be reanalysed (v) interpret the data as it links to planning, theory and external support (vi) provide plausible explanations for features of the data and disconfirm competing explanations (vii) create and describe new practices linked to the analyses and interpretation (viii) describe the theoretical framework on which this has been created, and (ix) demonstrate that the new practice can be practically implemented in the class without adversely affecting another practice (Lai, 2003).
analyses of within-school and cross-school achievement information. Eight schools have gone on to design and develop an electronic achievement management system. The idea is to electronically link schools' achievement management systems so that there is one bank of information with multiple entry points. Ideally, it could be used by teachers for planning lessons, by boards of trustees for school-wide planning and evaluation, by reform leaders to track district-wide achievement trends and design new interventions and by the Ministry and government to develop evidence-informed policy. It is a way of learning from one another to improve teaching, management and governance practices without unnecessary external accountabilities.

The process of writing the scenario confirmed my view that a more systemic approach to managing achievement information would make it easier to evaluate schooling improvement initiatives. It took a considerable amount of time to develop the system for eight schools. The school-based developers had to start the difficult task from scratch and the voluntary nature of the exercise led to several schools pulling out of the development. However, if their system is working in the near future and if it could be replicated quickly for lots of groups of schools, it is possible that the quality and use of New Zealand's schooling improvement evidence base may improve dramatically. That is not likely though unless the other reform groups are prepared to put themselves through the same intense learning programme as the principals and middle managers in Mangere. It also requires a commitment from the research and development community to provide services to lots of little groups of schools around the country. Given the limited preparedness of the research community to take up such a challenge, which was discussed earlier in this section, one has to wonder at the wisdom of attempting a widespread replication programme too quickly. It might be better to transfer the task to developers at the centre as a national strategy. A challenge in doing so is to retain the development of collegial accountability over unhelpful external accountability, which is so often associated with national achievement management systems.

4.2.3. Summary of the condition of the evidence in New Zealand

This section suggests that both the limited investments into programme evaluations and the diverse methodologies for those evaluations that have been completed contributed to the low quality evidence attached to schooling improvement initiatives
in New Zealand. Priority constraints explaining the low quality of the evidence include a policy silence on cross-school evaluations for schooling improvement initiatives, fierce professional opposition to centrally directed testing and limited systems for managing across-school achievement information. There is a possibility that a Ministry-supported replication of the achievement management system which is being developed in Mangere might help progress evaluation in a way that avoids external accountability systems that have caused numerous problems overseas. Evaluation policy supporting such a replication programme is not necessarily in conflict or competition with the other more established assessment policies, such as the formative assessment policy. That said, such a replication programme needs to be thought through very carefully so as to help build the overall capacity of the research community rather than tax the few within it who have expert knowledge and skills in programme evaluation. Without explicit policy guidance to help in this regard and to show how cross-school programme evaluation and within-school formative assessment can complement one another, there is an expectation that individual reform groups and boards of trustees in individual schools will continue to struggle with evaluating the success of their improvement interventions. That has proven to be a challenging task at which only a few groups have succeeded and that has taken the better part of a decade.

4.3. Summary and conclusions

New Zealand’s policy preference is for a developmental approach to programme evaluation that allows each initiative to create its own mix of methodologies. Although many of those methods have proven inadequate to date for evaluating the effectiveness of the initiative, the efforts have grown some local capacity in programme evaluation among some groups of schools. There are two apparent advantages associated with New Zealand’s slower and more distributed approach. One advantage is that as New Zealand’s evidence base gets stronger, schools’ governors and practitioners are not going to be as reliant on external experts to evaluate their efforts as their counterparts in the international initiatives. The second apparent advantage is that capacity in programme evaluation is being nurtured much closer to the classroom than it is in the international initiatives. The policy thinking
behind situating responsibility for programme evaluation in and around classrooms appears to be that useful findings from analyses of aggregated achievement information can be quickly transferred back into classrooms. These two advantages appear to be aimed at ensuring that teaching professionals do not become passive recipients of external experts’ opinions. Both assumptions imply that on-going reliance on expertise is not helpful. Yet, in other fields such as medicine, we willingly accept the need for on-going reliance on expertise so the policy logic here needs further thought.

Policy preferences in England and the United States have taken a different pathway by creating explicit sign-posts advertising the urgency of solving the underachievement problems. Those two countries have standardised an approach that relies on systemic achievement management systems. The contrast highlights two different types of thinking about the pace of change and accountability and the interrelationships between those two factors. The policy preference in England and the United States is an explicit sense of urgency and separating out development and evaluation accountabilities. Practitioners do the doing and external experts check on their effectiveness. The separation reflects a sense of urgency in finding out about the extent and nature of the achievement problem and the overall impact of the reforms on the problem. There has been no waiting for years for lead-practitioners to work out how to develop evidence of effectiveness. Evaluation experts have developed systems and produced achievement information from centres of expertise outside of schools. This policy preference does not assume the reform developers and evaluators do not talk to one another. To the contrary, connections between reformers and evaluators featured in the five international reforms analysed earlier in this thesis.

There are some cautions associated with the slower distributed approach. An obvious caution is that New Zealand may fall further behind other developed countries in producing high quality evidence. Another caution is that the capacity building exercise may be turning novice programme evaluators into slightly more skilled novices. The bulk of practitioners’ core work is about using appropriate content and pedagogical knowledge to construct, deliver and evaluate effective lessons. Achieving the right balance between development and research probably does not include turning practitioners into programme evaluators. While it may be beneficial to expect practitioners to become more knowledgeable about programme evaluation, the design and implementation of the evaluations should be left to expert
programme evaluators. Practitioners and expert evaluators can then form mutually beneficial connections to create strong planning, implementation and evaluation cycles within and across schools.

To conclude, the purpose of this chapter was to review the effectiveness of New Zealand's schooling improvement initiatives and select those with the strongest evidence for further analysis. The selection process was not an easy one but it did produce two initiatives for further analysis in Part III, namely the SEMO project and the Numeracy Development Project. An investigation into the overall condition of New Zealand's evidence base found that it was low in quality mainly because diverse methodologies are accepted and investment is geared towards development rather than evaluation. It is a long game in comparison to the centralised standards-setting exercises that took place in England and in the United States. A problem identified with the long game in relation to evaluating the effectiveness of schooling improvement initiatives is the considerable time it is taking school leaders to recognise that there is a problem to solve. It is also inadvertently dismissing the useful contributions expert programme evaluators could make to lift the quality of the evidence base to a much higher level. The issues and opportunities raised here in Part II are re litigated in the more theoretical discussion in Part IV, the final part to the thesis. The main aim of that part of the discussion is to help ensure that future developments in programme evaluation of schooling improvement initiatives in New Zealand do not widen the gap between the quality of our evidence base and that of other developed countries like England and the United States.
PART III: Development and Implementation Characteristics

PART III
Development and Implementation Characteristics

Purpose

The purpose of Part III is to conduct a series of investigations to identify, describe and explain development and implementation characteristics attached to the seven initiatives selected for analysis that are relevant to accelerating the learning of effective reform practices. The investigations provide the findings necessary to answer the third research question: What can be learned about developing and implementing effective school improvement from the national and international examples selected for analysis?

Approach

Part III involves two investigations into reform learning processes; (i) development and implementation characteristics of the seven initiatives relevant to learning effective reform practices (Chapter 5 & 6) and, (ii) policy-research-practice collaborations featured in the model for the initiatives operating in New Zealand (Chapter 7). These two investigations reflect a critical inquiry that starts broadly and then gets into some depth in areas that seem most useful to accelerate reform learning in New Zealand’s schooling improvement movement. The investigations presented
PART III: *Development and Implementation Characteristics*

two major methodological challenges. The first challenge was to create some order in sifting through and extracting out relevant information from the considerable number of relevant research, evaluation and policy documents. My solution for this challenge was to create several hard copy file boxes and electronic files to store relevant documents in such a way that they were quickly accessible. In the first instance, the titles of the boxes tended to be generic, such as ‘schooling improvement international’, ‘schooling improvement New Zealand’ and ‘assessment’. As my thinking progressed and the focus of the inquiry sharpened, the storage arrangements were expanded to include file boxes specific to the initiatives being considered for analysis. I arranged the boxes in my study in rows on shelves so that I could easily dip into any one of them as the need arose. That was important as the tightening process meant re-visiting the same documents many times to check for accuracy or to find something slightly different from previous searches.

The second major methodological challenge was to come up with some analytical frameworks that would help transform the vast amount of source documentation into something that would adequately answer the research question at hand and would be both interesting and useful to the reader. As things panned out, two such frameworks were developed, one for each investigation. The first one is referred to as the comparative framework (Figure 5). A brief outline of its development is included in the paragraphs immediately below. It explains methodological issues specific to the first investigation. Because developing this first framework was such a major undertaking and so important in sharpening the focus of the other investigation, an entire chapter (Chapter Five) is dedicated to presenting and explaining it more fully than was possible in this introductory section of Part III. The other analytical framework helped to get some depth to the second investigation. That framework, called a framework for analysing policy-research-practice collaborations (Figure 4), is a tool to analyse collaboration among policy developers, researchers and practitioners. Methodological issues relevant to this framework and the second investigation in general follow the explanation of first investigation below.

Investigation One begins with a four-step process to develop the comparative framework and then uses the framework to construct a series of learning process models for the seven initiatives selected for analysis. The first step in the process of developing the framework involved synthesising available information about the
seven initiatives and creating succinct descriptions of them for analysis. Succinct descriptions of the initiatives are provided in Appendix B. A template (Table 9) was developed to collate descriptive and evaluative notes taken from primary documents and associated research for each initiative.

Table 9. Template Table for Taking Descriptive and Evaluative Notes of the Seven Initiatives Selected for Analysis.

<table>
<thead>
<tr>
<th>(i)</th>
<th>Type of initiative</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Developer:</td>
</tr>
<tr>
<td>b.</td>
<td>Type and scale:</td>
</tr>
<tr>
<td>c.</td>
<td>The problem</td>
</tr>
<tr>
<td>d.</td>
<td>The design theory</td>
</tr>
</tbody>
</table>

(ii) Reform practices

(iii) Support infrastructure

(iv) Timeframe for significantly improving student achievement

(v) References

Italicised descriptor headings helped organise the information so that it could be analysed thoroughly in the construction of the comparative framework. Each table has the title of the initiative as the heading and five content sections.

(i) The developer, type and scale of the initiative as well as the problem that the initiative is trying to solve and design theory for the intervention approach. These elements help locate where the reform fits within the broader picture of schooling improvement. They also help understand the reasons that ruled in a particular approach and ruled out other possibilities.
PART III: Development and Implementation Characteristics

(ii) A list of the practices outline what the practitioners were expected to learn.

(iii) The support infrastructure explains how the practitioners were expected to learn those practices.

(iv) The timeframe helps to understand how long it took to make a positive impact on student achievement.

(v) List of reference documents used to create the description.

Table 10 provides a pen portrait of the five international and two national initiatives for readers who do not wish to dwell on the detail of the descriptions in Appendix B. The first one is situated in England, the next four in the United States and the last two in New Zealand. The form and scale of the seven initiatives varies. England’s Literacy and Numeracy strategies are national professional development programmes. They are the largest in scale spanning all 22,000 primary schools across the country. Of the initiatives in the United States, Direct Instruction, Success For All and The School Development Programme are comprehensive school reforms operating in hundreds or thousands of schools whereas New York District #2 was a smaller-scale district-wide reform intervening into 32 schools. New Zealand’s Numeracy Development Project is a national professional development programme but in a much smaller country with just under 2000 schools involved. The SEMO project is another district-wide reform like New York District #2, but with some interesting differences. Of particular interest in the investigations in Part III are the learning networks that formed through a relationship-building exercise early in the reform process. A network is simply a connection between entities, such as people or electronic nodes (Barabasi, 2002). A learning network is defined as a connected group of reformers working together non-hierarchically to find out what and how to solve achievement problems among disadvantage students.
PART III: Development and Implementation Characteristics

Table 10. Pen Portraits of the Seven Initiatives Selected for Analysis.

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>England’s national literacy and numeracy initiatives</strong></td>
<td>A national reform led by a lead bureaucrat, Michael Barber and directors of newly established national literacy and numeracy centres to address long-term neglect of disadvantaged students’ entitlement to quality schooling. The reforms, which were mandated for all 22,000 primary schools across England, introduced standardised frameworks for literacy and numeracy as well as scripted lessons and resources. The support infrastructure in this case represented a unification of the bureaucracy at the national, regional and local levels with the help of 300 consultants to help introduce and monitor the standardised practices in the schools. Positive overall results were achieved in about three years.</td>
</tr>
<tr>
<td><strong>Direct Instruction</strong></td>
<td>A current Comprehensive School Reform designed by Siegfried Engelmann at the University of Oregon to address the problem of ineffective teaching. The programme which operates in approximately 300 schools across the United States involves highly scripted reading, language and mathematics lessons and standardised resources for Grades 1 and 2 students and whole class instruction beyond Grade 2. Reviews of progress are conducted weekly and a corrective reading programme helps catch up those Grade 4 students who are seriously behind. A support infrastructure in the form of a National Institute of Direct Instruction, institute consultants and a school-based management team tends to help schools achieve positive results in two to three years.</td>
</tr>
<tr>
<td><strong>Success For All</strong></td>
<td>A current Comprehensive School Reform designed by Robert Slavin and Nancy Madden at John Hopkins University to replace ineffective remedial programmes. The programme operates in approximately 1300 schools across the United States as well as in many schools in other countries. The programme groups students into classes with similar reading levels and uses proven mastery and collaborative teaching and learning techniques to teach students how to read. A family support team is also set up to encourage reading at home and progress is monitored through eight-weekly assessments. The support infrastructure includes a headquarters in Baltimore with consultants overseeing in-school facilitators to set up and sustain the programme. Positive results are usually made in two to three years.</td>
</tr>
<tr>
<td><strong>The School Development Programme</strong></td>
<td>A current Comprehensive School Reform designed by Dr James Comer at Yale University to address any developmental needs of students at entry to school. The actual number of schools participating in the programme is unspecified in relevant literature but it is believed to be in the many thousands. The programme involves three teams (a school planning and management team, a student support team and a parent team) concentrating on improving school climate dimensions and creating caring student-teacher relationships and fair and equitable discipline. Infrastructural supports include several development units at the Yale Child Study Centre and implementation co-ordinators helping the school-based teams get positive results, which usually take three to four years.</td>
</tr>
<tr>
<td><strong>New York District #2</strong></td>
<td>A district-wide reform led by the Department of Education’s district office superintendent and team to help the 32 publicly-funded schools in the district deal with challenging national achievement standards from 1987-1998. The reform centred on individual school improvement plans negotiated between the district office officials and the schools’ principals. Close monitoring of progress involved regular on-site visits and analyses of city-wide test information. Infrastructural supports included the district office working in partnership with a research team to agree on priorities for development and contracted consultants to help principals and teachers implement the improvement plans. Results gradually improved over the decade that the reforms were in place.</td>
</tr>
<tr>
<td><strong>The Numeracy Development Project</strong></td>
<td>A current national professional development programme in New Zealand which was designed in 1997 and iteratively developed since by a collaboration of policy developers, researchers, developers, publishers and lead</td>
</tr>
</tbody>
</table>
The programme aimed to address the New Zealand students' overall low performance in solving number problems in the 1995 Third International Mathematics and Science Study (TIMSS). The programme includes a framework of typical student learning in number knowledge and a professional development programme for teachers of Year 1-10 to learn to conduct regular diagnostic assessments and to plan and implement knowledge and strategy lessons in number. Infrastructural supports include a collaborative design group, a formative research team and professional development facilitators. Positive results were reported about five years into the reform effort.

The SEMO project was a district-wide reform intended to Strengthen Education in the two districts of Mangere and Otara, hence the name SEMO. The project was announced in 1997 by the Minister of Education, Wyatt Creech, as a partnership initiative between the Ministry, the 42 schools in the two districts and the community leaders. Several years were taken up resolving relationship problems and improving governance and management systems in schools. The project then focused on developing learning networks among local officials, researcher-developers and practitioners. The learning networks designed, implemented and evaluated a series of interventions into classrooms, which helped create a strong discipline around planning and assessment practices. Although some networks achieved positive results within six months, overall improvement has been a gradual process over almost a decade.

The second step in developing the comparative framework involved critical analysis of the descriptions of the initiatives in order to identify common conditions related to learning effective reform practices. Five common conditions were found; strong designs, standardised practices, gaining acceptance in schools, learning supports and systematic monitoring of the students' achievements. These conditions were found by searching through source documents and grouping and re-grouping information until it made sense in terms of explaining what and how reform practices were expected to be learned. Several searches of the source documents were completed to see if there were any alternative groupings that might be more appropriate. Additionally, I checked the school improvement literature and several related bodies of literature to see if there were relevant groupings I had overlooked. Four reviews were found that synthesised a considerable amount of relevant research (Borman et al., 2003; Harker, 2002; Harris, 2000; Potter & Reynolds, 2002). Although the groupings in those reviews described elements of successful school reform more broadly, they were useful for fine-tuning my groupings relevant to learning effective school improvement practices. Each condition is explained and elaborated by referring to evidence in source documents that were used to create the

---

28 Important themes from the four reviews included: a focus on student outcomes; professional capacity building particularly in learning effective instructional practices; high expectations; taking account of local circumstances; resources directed to classroom improvement; high quality design; fidelity to programme implementation; and, outcomes-focused evaluations.
descriptions of the initiatives in Appendix B. Comment is also made about the relevance of each condition to practitioner learning.

The third step in developing the framework was to ensure that descriptions of barriers to learning were built into it. Descriptions and analysis of common conditions of successful international school reforms alone present what one group of researchers refer to as “sugar-coated” descriptions (Rowan et al., 2004). In other words, presenting findings about learning supports for practitioners without also outlining the barriers to learning that practitioners typically encounter is not an accurate reflection of what is happening. In order to create a balanced picture, there is a need to describe examples of barriers to learning and how they were managed to minimise interference within the five international reforms. In that regard, the available information was more implicit than explicit. A broader literature review of studies that have identified barriers to practitioner learning processes were used as a proxy to get started (Borman et al., 2003; Rowan et al., 2004; Spillane et al., 2002; Timperley & Annan, 2004; Timperley, Phillips et al., 2003; Vogel, 2003). Then examples were found in the five initiatives of management strategies to complement that information.

The fourth and final step was to transform the relevant findings into a useful framework. I decided to use the model presented in Figure 2 (p.25), which situated the macro policy contexts in relation to schooling improvement initiatives, as the foundation of the framework. What was particularly useful about that model was the use of concentric circles to represent contexts and the use of arrows to represent influences across the contexts. Because the comparative framework was intended as a tool to help understand the learning processes set up for practitioners in the initiatives, it needed to have learning dimensions and barriers built into it. Figure 2 did not have such dimensions. That is where the analysis of the five common conditions and the barriers to learning came into play. An explanation of the process by which the five common conditions were transformed into learning dimensions is included in the description of the comparative framework in Chapter Five.

The decision to analyse and compare learning dimensions is consistent with the view among some researchers that learning effective school improvement practices and implementing them is a distributed responsibility across the schooling system (Coburn, 2005; Firestone, 2005; Fullan, 2005b; D. Hargreaves, 2003a; Spillane et al.,
PART III: Development and Implementation Characteristics

2002). Spillane et al (2002), for instance, argue that sense-making in relation to standards-based reform policy directives in the United States is as much about the thinking and actions of school leaders, district administrators and researchers as it is about teachers' thoughts and acts in implementing the reform practices. They draw on situated and distributed cognition theory (Greeno, 1998 in Spillane et al., 2002) to argue that "sense-making and action are distributed in the interactive web of actors, artefacts, and situation, and this system becomes the appropriate level of analysis" (p.404). In support of that argument, the development of learning dimensions as a system for analysis helps to avoid the development of a new list of conditions which assume that practitioners can use such lists to solve achievement problems.

With the comparative framework in hand, it was possible to construct models that represent the processes developed in the seven initiatives for practitioners to learn effective reform practices. Three such models were constructed. The reason for developing the three models was to compare the learning processes used in the initiatives in England and the United States with those used in New Zealand, a much smaller country that has not tried to replicate reform processes typically developed in those larger ones.

The models were developed through an iterative review and revision process. The iterative process meant the models were revisited many times. It was a matter of constantly adjusting the component parts of the models until there was a best fit with the published accounts of how learning processes were developed and implemented in initiatives. It also meant re-visiting regularly the massive amount of publicly-available information that I was using as an evidence base. That did not mean re-checking the same idea time and time again. Rather, each revisit was undertaken with slightly different thoughts from the time before. What I was looking for in each successive search had often been overlooked in earlier searches. The models are my own constructs and do not in any way attempt to represent the collective views of the policy developers, researchers and practitioners who commented along the way.

Once initial models were created, the descriptions of the initiatives and the common conditions and barriers were reanalysed several times in order to find disconfirming evidence for the way the models were constructed. That task involved critical feedback from policy developers, researchers and practitioners involved in the
PART III: Development and Implementation Characteristics

field of school improvement within New Zealand and overseas. That feedback task was important in case the broad level of aggregation distorted the way policy makers, researcher-developers and practitioners approached school improvement in the different countries. Discussion about the New Zealand-based model was much easier to generate than it was for the international models. As my work is exclusively in the field of school improvement within New Zealand, it was possible to share the model with a wide range of policy developers, researchers and practitioners.

The latter part of the investigation in Chapter Six is dedicated to an analysis of the learning connections occurring between the policy developers, researchers and practitioners in the seven initiatives. What became apparent from the models was the importance of relevant knowledge flowing between participants from the policy, research and development and practice communities. Of utmost importance in the knowledge flow was that practitioners came into contact with knowledge relevant to solving their students' achievement problems, understood it and accepted it as useful and then used it to improve their practice. The flow of knowledge was a process of acquiring relevant knowledge and transforming it into practice. Learning connections were a means of making the process work effectively. The vertical and horizontal arrows within the three models provide a general guide about how the transformational process happened in the seven initiatives. They allude to different sorts of learning connections that help manage the transformation process.

Investigation Two. The main reason for focusing solely on the initiatives in New Zealand in this inquiry was because such collaborations were not found in the English and United States models. Rather, policy developers tended to remain at arms-length from the action and researcher-developer/consultant-practitioner collaborations (referred to as research-practice collaborations) tended to be the norm as distinct from

---

29 Feedback on the international models involved regular doctoral supervision feedback from Professor Viviane Robinson and Associate Professor Roger Dale and a series of presentations with follow-up critical feedback. Some of the more influential feedback came from the following presentations: (i) Ministry of Education Managers seminar (16 February 2005) (ii) A discussion with 12 Chiefs of States from the United States (3 March 2005) (iii) Professor Bill Firestone (21 March 2005) (iv) Ministry of Education senior manager feedback (Adrienne Alton-Lee and Kathy Smith) (March & early April 2005) (v) Ministry of Education Curriculum: Teaching and Learning team seminar (April 5, 2005) (vi) Audience feedback at session 31.67 of the American Education Research Association conference in Montreal (12 April 2005). The audience included United States and English educators. Michael Fullan was also present and provided useful feedback as one of research team which evaluated the English Literacy and Numeracy reforms and has commented extensively on reform developments in the United States (M. Fullan, personal communication, April 12, 2005).
policy-research-practice collaborations. The two initiatives in New Zealand featured numerous collaborative interactions between policy developers, researcher-developers and practitioners.

Stein and Coburn's (2005) conceptual framework for analysing effective interactions between the research and practice communities was found in the literature to be the most useful tool for analysing those interactions. Their framework (Figure 3) places the research community alongside the practice community with semiautonomous trajectories. Both trajectories begin with past understandings and practices which feed into the present improvement tasks and end with changes being made to the original understandings and practices. They are only semiautonomous because there is interaction between the two trajectories, referred to by Stein and Coburn as the “third space” and what happens within that space influences both the research community and the practice community. It is within the third space that researchers and practitioners were found to connect with one another in relation to four improvement tasks: design of research, designing for scale, accounts of educational improvement and systems for developing, codifying and sharing of practitioner knowledge. The latter two tasks are particularly relevant to the analysis of the policy-research-practice collaborations in New Zealand initiatives.

Figure 3. Stein and Coburn's (2005, p.8) Framework For Analysing Research-Practice Collaborations.
PART III: Development and Implementation Characteristics

Community of practice theory (Lave & Wenger, 1991) is used to explain the nature of the relationship between the researchers and practitioners and what happens within the third space. Researchers belong to various communities of practice within and across tertiary organisations such as universities. They are researcher communities of practice and they deal with things that are important to them within the tertiary world. Likewise, practitioners belong to communities of practice within and across schools. They are practitioner communities of practice and they deal with things that are important in their schooling world. When researchers and practitioners come together in the third space they form new communities of practice, commonly referred to as research-practice collaborations. Stein and Coburn (2005) argue that they are particularly useful for developing domain knowledge and tools that practitioners can use to become more effective. In the framework, they are referred to as ‘boundary practices and objects’. Research-practice collaboration is also a useful way of helping researchers learn alternative ways of influencing teachers to change other than simply writing and publishing reports.

Third space communities of practice break the mould of practitioners being the recipients of researchers’ findings. Practitioners instead are reform co constructors alongside researchers. Initial connections between the two groups tend to help practitioners make sense of knowledge transfers to the point that they can make a start. Ongoing interactions help them acquire a deeper knowledge and sort out implementation problems as they arise. Similarly, on-going connections help researchers understand practitioners’ successes and difficulties in trying to make changes in their classrooms. Researchers are, in turn, challenged to help the practitioners deal with the difficulties and achieve further successes. Therefore, critical to the formation and development of this particular sort of community of practice is the “regular, ongoing and practicable” connections between researchers and practitioners (Stein & Coburn, 2005, p.8).

A limitation of Stein and Coburn’s (2005) framework for the analysis in this chapter is that it concentrates solely on third space interactions between researchers and practitioners. It does not account for the involvement of policy developers in the learning process. Yet they feature strongly in the way learning processes for practitioners have been developed and implemented in the Numeracy Development Project and the SEMO-related initiatives. Consequently, Stein and Coburn’s (2005)
framework needed to be adapted to be useful in analysing those two initiatives. The adapted framework is presented in Figure 4.

Figure 4. A Framework for Analysing Policy-Research-Practice Collaborations
Adapted from Stein and Coburn’s (2005) Framework.

There are two additions to the original framework. The first addition is another trajectory for the policy community. It is situated in the centre of the framework. The national policy community is intentionally positioned below the research and school communities in the diagram. The reason for that placement is because of the New Zealand policy preference to promote leadership of school improvement initiatives within the research and practice communities and not within the policy community. The preference is for policy to influence practitioners as an underlying support mechanism rather than to direct them through rule-governed policies from above (Annan, 1999). The second addition is a fourth space sitting below and behind the third space. It is the space within which policy developers, researcher-developers and
lead practitioners worked together on aspects of the reforms. That space acknowledges that while some things were developed through research-practice collaborations, other developments occurred through policy-research-practice collaborations. The existence of the latter type of collaboration is a crucial point of comparison to what was happening in England and the United States.

Attention now turns to the first investigation which provides broad analytical findings into the learning processes used in the initiatives in England, the United States and New Zealand.
PART III: Development and Implementation Characteristics
CHAPTER Five: A Comparative Framework

A COMPARATIVE FRAMEWORK

The first half of this chapter presents the common conditions and barriers found in the five international schooling improvement initiatives in order to design the comparative framework and create learning process models in the next chapter. In the second half of the chapter, the comparative framework is introduced with accompanying explanations of how its component parts were constructed.

5.1. Common Characteristics of the Five International Initiatives

Four common conditions were found which were considered relevant to accelerating the learning of effective reform practices: a high degree of standardisation in non-negotiable design theories and implementation practices; developers gaining schools’ acceptance of their preferred theories and practices; learning supports and evaluative thinking. The four conditions are discussed in turn before revealing the common barriers that tend to interfere in reform learning processes.

5.1.1. Standardisation

Standardised design theories and implementation practices was the first and most dominant common condition that the analysis revealed. “Standardised” theories and practices are those that developers external to schools create and expect practitioners to implement uniformly across different contexts. All five initiatives had explicit and not-negotiable design theories (See details in Appendix B). An analysis of source documents that contained information relevant to those theories indicated an overarching theory that we should aim at students from disadvantaged communities learning at expected levels (Adams & Engelmann, 1996; Comer, 2005a; Earl et al., 2003; Elmore & Burney, 1998; Slavin et al., 1992). The analysis led me to believe that there were at least three underlying assumptions attached to the overarching theory.

The first assumption inferred from source documents is that the students have to be taught effectively for them to learn. Direct Instruction, for instance, assumed students
can learn at an age-appropriate rate by using mastery learning techniques at a young age and, as they get older, by applying their mastery skills and accumulated knowledge to new situations (Adams & Engelmann, 1996). The second key assumption is that learning amongst students from disadvantaged communities has to be accelerated. The speeding up process was about students catching up to the norm expected of their age group. In England, for instance, the catch-up was expected by age eleven when all students should be at level four for reading and mathematics (Barber, 2000a). A third key assumption is that effective teaching in classrooms, on its own, is an inadequate solution. The behaviours of other adult groups, such as district administrators, principals and parents, were targeted for improvement. The School Development Programme leaned most heavily on this assumption (Comer, 1996). Combined, the three assumptions represented an overwhelmingly optimistic view about solving underachievement problems among disadvantaged students.

There were two distinctly different methods for developing the design theories. One way was to use existing theories. The developers of the three comprehensive school reforms did this. Each developer used their extensive research experience to select learning principles that would endure over time in an environment of constant change (Comer, 1996; Engelmann, 1999; Slavin & Fashola, 1998). Engelmann selected mastery learning while Slavin preferred collaborative learning and Comer wanted to create a balanced ecological environment surrounding children's learning. Each of these learning principles has had high credibility in the educational and psychological research communities for a considerable period of time (Bloom, 1971; Bronfenbrenner, 1979; Idol, Nevin, & Paolucci-Whitcomb, 2000).

The other method was to develop the theory along the way. This formative method was most prominent in the New York District #2 reforms. The superintendent got something going in the form of a district-wide instructional improvement plan and contracted researchers to inform the plan and check for causal links between the reforms and student achievement trends (Elmore & Burney, 1997, 1998, 2000; Harwell et al., 2000). In regard to design theory, the researchers referred to the method as continuous improvement. The approach taken to design the literacy and numeracy initiatives in England was both pre-determined and formative. No incremental changes were permitted in the early years of the reforms. However, when results were not up to expected levels after three years, the national directors
considered changes to the design theory and the standard practices. Critical feedback from a contracted research team (Earl et al., 2003) and non-contracted researchers (Brown et al., 2000) appeared to help influence the change in direction.

In terms of implementation practices, each initiative had a distinctive set of standardised practices rather than there being one universal set. Three initiatives (Direct Instruction, Success For All and the English reforms) had prescribed pedagogical practices and curriculum resources that left no doubt about what would be taught in classrooms and how school-wide systems would support teachers. In New York District #2 there were standard procedures around negotiating achievement standards and improvement plans. Principals and teachers were expected to deliver on the plan and if they did not they faced losing their jobs (Elmore & Burney, 1998). In contrast, implementation practices in the School Development Programme revolved around three collaborative teams that had standard functions and rules. Although the standard practices varied markedly across the initiatives, the developers expected high fidelity to programme implementation within each initiative.

Because classroom and/or administrative practices were standardised, practitioners only had to concern themselves with implementation issues. Principals and teachers did not have to worry about design and programme evaluation issues because developers took responsibility for those sorts of things in centres of expertise external to schools. However, the practitioners were not passive as they were required to fit the reform practices into their various settings. Care was taken by the co-ordinators to help the practitioners do this. They helped adjust systems and practices surrounding the standardised practices, such as altering strategic plans to prioritise achievement targets and implementation procedures associated with the reforms (Direct Instruction, The School Development Programme, and New York District #2). They also helped practitioners let go of personally-valued alternative theories and practices. This type of contextualising did not allow the practitioners to alter the standardised practices.

5.1.2. Gaining acceptance

The second common condition relevant to how practitioners were expected to learn is about the way developers went about gaining practitioners' acceptance of the standardised theories and practices. Acceptance was gained through two different
methods. The first method was a mix of educational expertise and funding support that caused schools to forfeit their autonomy by choice and commit to the standardised features of the initiatives. This method was used by the three comprehensive school reform developers, Engelmann, Slavin and Comer. Their extensive educational knowledge, experience and contacts gave them credibility in the education community. They backed up this credibility with indirect positional power gained by assisting schools and school districts to access considerable additional federal, State and city funding. Success For All’s developers, for instance, did this by matching the goals of their programmes to the goals of various funding pools. As criteria for the funding pools evolved, such as the federal funding criteria for Title 1 funding, Slavin altered the goals of the programmes to match (Success For All, 2003).

The second method of gaining acceptance among practitioners was to introduce the standardised practices through top-down mandates with the backing of experienced and credible educators. Department of Education officials in New York District Office #2 and central government in England used their positional authority to identify problems they considered priorities and to mandate standard reforms to solve them. They ruled in uniformity and ruled out autonomy from the outset. They did not, however, rely solely on positional authority to gain acceptance. Superintendent Alvarado and his two lead administrators, for instance, had extensive experience in education and developed a formative relationship with researchers attached to the University of Pittsburgh (Elmore & Burney, 1997, 1998, 2000; Harwell et al., 2000). This gave the district office officials considerable knowledge-based authority.

Practitioners accepted what experts external to their schools told them to do. Principals and teachers who thought they knew better or drifted from the prescribed practices tended to be treated as a threat and sanctions were imposed. A relatively subtle sanction was the National Institute of Direct Instruction’s policy not to support schools that did not commit fully to the programme (K. Engelmann, personal communication, 2004). A more overt sanction was the replacement of teachers and principals in New York District #2 for not adhering to the district’s instructional improvement plan (Elmore & Burney, 1998). Together, the standardised approaches and sanctions meant that effectiveness was given more weight than teacher and principal autonomy.
5.1.3. Vertical and horizontal learning

The third common pattern which helps to explain how the practitioners went about learning was the development of multi-layered support infrastructures which operated in two learning dimensions referred to as the vertical and horizontal learning dimensions. In a nutshell, the vertical dimension promoted hierarchical learning through formal structures in the education system and the horizontal dimension promoted non-hierarchical learning within and between formal structures. I found these two dimensions after sifting through the source documents several times and then checking to see if there was anything written about these arrangements in the general schooling improvement literature. There were a few references for vertical and horizontal (or lateral) learning in the schooling improvement literature (Fullan, 2005a, 2005b; D. Hargreaves, 2003a, 2003b; Hill, 2001; Veugelers, 2005) but none of those explored both dimensions simultaneously in any detail. The two dimensions helped me to make sense of how professional capacity building was being handled in the international initiatives. They also helped me to ascertain the extent to which local circumstances were being taken into account. A little more detail about the two learning dimensions is useful here to foreground the learning process models presented in the next chapter.

The vertical learning dimension involved those with positional authority setting up top-down learning mechanisms that transferred desired practices and knowledge into the schools and classrooms. Earle et al.’s (2003) descriptors for the layers of support infrastructure for England’s literacy and numeracy initiatives are useful for describing three layers of the vertical learning dimension common to the five initiatives. A common top layer of support was a ‘centre of expertise’ where developers external to schools designed and governed the initiative. A middle layer involved ‘bridging’ arrangements which created feedback and feed-forward loops between the centres of expertise and schools. At the bottom, in-school supports assisted practitioners to implement the practices intended by the developers to improve school-wide systems, classroom practices and, in the School Development Programme, home practices. The multiple layers of support ensured that practitioners were not left to their own devices to learn and apply the developer’s preferred theories and practices. The vertical dimension also had a bottom-up component. Findings about what was
happening in the classrooms were channelled back up through the hierarchy to inform the development of the initiatives. For example, England’s numeracy directors reconsidered the use of whole-class teaching as a blanket reform approach when a group of researchers found that it had a negative achievement effect on the lowest performing students (Brown et al., 2000).

The horizontal learning dimension manifested itself through various grouping arrangements that encouraged collegial learning. Conference calls, coaching groups and principals’ forums and academies were deliberate attempts to get practitioners to help each other deal with implementation issues as they arose. Those types of activities were ways of building communities of interest around the reform practices. This concept, commonly referred to as “communities of practice” in the literature (Wenger, McDermott, & Snyder, 2004), has been recognised as a feature of the business world for the past two decades. Wenger et al (2004) define communities of practice as: “groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis” (p.4). Although the concept is relatively new to the school improvement field, horizontal learning through communities of practice is recognised as a useful way for practitioners to collaboratively make sense of the reform practices and goals that they are expected to implement (D. Hargreaves, 2003a; Spillane et al., 2002).

The vertical and horizontal supports acknowledge that practitioners can learn as much from informal situations as they can from formal professional development programmes. While the vertical dimension tended to support deliberate and planned learning, the horizontal dimension gave participants opportunities to engage in spontaneous learning by discussing implementation issues with colleagues. Intentionally developing deliberate and non-deliberate learning opportunities is a powerful way to help practitioners match intentions to outcomes and detect and correct errors. Both of these tasks are recognised as critical elements of organisational learning relevant to schooling (Argyris, 1999; Robinson, 2002).

Organisational learning was not the only type of learning that was important within the five initiatives. There appeared to be context-specific learning agendas. In the centres of expertise, for instance, the reform developers were centred on design and intervention learning. Meanwhile, teachers and students in classrooms were focused on academic teaching and learning and parents (in the School Development
Programme) were encouraged to focus on learning how to contribute to their children’s academic learning. These context-specific learning agendas were linked by the feedback and feed forward loops created by the co-ordinator roles. Together, the context-specific learning agendas and the learning links between the contexts appeared to help create a connected learning system.

5.1.4. **Strong Evaluative Thinking**

The fourth and final common condition, which also related to how practitioners were expected to learn, was the way that they were encouraged to evaluate what was going on in their classrooms. They were encouraged to concentrate on formative assessment and to leave the task of programme evaluation to the developers. That separation of evaluative functions meant principals and senior managers scrutinised achievement information to check that their management systems were effective and, with their teachers, to check that the teaching practices (or climate dimensions in the School Development Programme) were having a positive impact on student achievement. At the same time, the developers analysed achievement information to review and revise the design of the reform programmes. Engelmann, Slavin and Comer tapped into their own extensive research expertise to do this while Alvarado in New York District #2 and Barber and more latterly Hopkins in England tapped into the research community to help them out (Earl et al., 2000; Earl et al., 2003; Elmore & Burney, 1997, 1998, 1999, 2000; Harwell et al., 2000).

My analysis revealed that the co-ordinators (also referred to as consultants in New York District #2 and England), working in schools on behalf of the initiatives, acted as connectors between the two functions of programme evaluation and formative assessment. The term “connector” is used in network literature for identifying people or electronic devices that create connections between various groups of people and information banks in complex systems (Barabasi, 2002). In the five initiatives, the co-ordinators went into the schools on behalf of the centres of expertise and helped collate achievement information for both programme evaluation and formative assessment purposes. They participated with practitioners in regular cycles of monitoring student achievement information. The time interval

---

30 The various co-ordination roles were additional to the standard staffing schedules of schools.
between monitoring sessions varied. Direct Instruction's cycle, for instance, was every week (K. Engelmann, personal communication, 8 January, 2004) whereas Success For All’s cycle was eight weekly (Success For All, 2003). The co-ordinators also forwarded relevant information about achievement trends and process issues to the centres of expertise to constantly review and, if necessary, revise the design of the initiatives.

Achievement information was not a taken-for-granted commodity. The developers of all five initiatives set up or tapped into achievement management systems that collected and collated achievement information. England’s literacy and numeracy directors set up a national testing system to monitor the effectiveness of the initiatives across all 22,000 primary schools. The developers of the three comprehensive school reforms also set up dedicated achievement information management systems in their centres of expertise. District office officials in New York District #2 tapped into existing state and city-wide achievement databases to monitor the impact of the schools’ improvement plans. The key point here is that achievement information was readily available as a primary indicator of success.

These evaluative practices meant practitioners could be held accountable for implementing the practices effectively. The combined effect of the programme evaluations, research projects and formative assessments was to de-privatise the achievement problems and the degree of progress being made in solving them. There was little chance of escaping acknowledgement for success or scrutiny for failure. What was interesting was that the practitioners were not holding each other to account for progress against aims. They were accountable to the designers of the initiatives. Although the external accountability made sure that the principle of programme integrity was upheld, it did not encourage collegial accountability.

5.1.5. **Summary of common conditions**

In summary, a great deal of effort was put into making school practitioners’ lives manageable. They were mainly told what to do and then closely monitored in the way they went about their business. Consequently, practitioners’ involvement in planning and evaluation was down-played although not to the point that they had no involvement at all. Some researchers are critical of standardised approaches across schools because they tend to disregard the highly contextual nature of schooling.
(Carpenter, 2000; Eisner, 1992; Kliebard, 2002). The criticisms are based on an argument that trying to standardise what happens in a classroom takes away a teacher's responsibility and ownership of the specific learning needs of her or his students. Consequently, teachers can go through the motions like factory workers on a product line. Those criticisms are generally fair as most standardised reforms have failed to improve achievement among disadvantaged students (Fuhrman, 2002; Kliebard, 2002). However, all five initiatives in this study are exceptions to the criticisms because they had a significant impact on student achievement within about three years.

5.2. Common Barriers to Practitioner Learning

A number of barriers that can impede practitioner learning have been identified in numerous studies. One group of researchers tracking the development of comprehensive school reforms in the United States identified four common barriers; a highly politicised context; inauthentic teacher buy-in mainly due to power relationships with administrators; evaluative information not routinely informing planning; and a lack of adequate resources including teacher time (Rowan et al., 2004). Another researcher found that government officials can be a barrier to practitioner learning when their influence is based on piecemeal understandings and a lack of critical instructional details (Spillane, 2000). Other researchers have found that the cognitive capacity and emotional state of the practitioners can be barriers to their own learning, such as: beliefs that what they are expected to do is unreasonable (Borman & Hewes, 2002); debilitating emotions such as anxiety, confusion and frustration (Spillane et al., 2002); inadequate requisite cognitive skills, assessment literacy and pedagogical content knowledge (Spillane et al., 2002; Timperley, 2004a; Vogel, 2003); and getting distracted by less important issues, such as building upgrades (Timperley et al., 1999), and by unnecessary tasks like teaching to tests (Datnow & Stringfield, 2000).

These barriers appeared to be managed through the application of strong design theories and through vertical and horizontal learning mechanisms. Strong designs helped deal in a purposeful way with schooling improvement's complexities, frustrations, set backs and distractions. Design theories are inadequate, however, if they are not applied effectively. Often theories to solve the problem of
underachievement are espoused but the associated practices come up well short of the desired results. Kliebard (2002) usefully refers to the theory-practice gap in schooling improvement as a rhetoric of improvement. Lofty espousals that are not helpful in overcoming reform difficulties or useful ones that are ineffectively applied in practice fit this reference. The reference is not true of the five initiatives analysed in this study. Substantial evidence is presented (Appendix 1 & Chapters 2 & 3) that links the theories to the desired practices and the positive achievement outcomes. The success meant there was a credible relationship rather than a gap between the theory and practice of improvement.

Vertical learning supports tended to manage the more overt barriers, such as the adequacy of resources to implement the reforms. Because of the dominance of the vertical dimension in the five international initiatives, there appeared to be little opportunity for such barriers to form and hinder practitioner learning. In the English reforms, for instance, the adequacy of resource allocation was tested early on. Teacher unions raised concerns about insufficient funding for training and development in the early stages (Earl et al., 2003). The Government immediately provided additional funding. An immediate response in this case was particularly important to avoid the potential for considerable teacher dissent because of the high levels of prescription, work load issues and exposure to public criticism that the developers were expecting the practitioners to live with.

Managing tacit barriers, such as teacher beliefs and emotions, appeared to be left to the horizontal supports to sort out, such as mentoring strategies, conference calls, the principals’ forums and the like. Sense-making opportunities such as these are critical because tacit barriers are much more difficult to manage than overt ones. They often remain unseen and unheard. Whether practitioners buy into tacit barriers or not is associated with their cognitive filtering processes which rule in and rule out what they believe in. Cognitive filtering is commonly associated with schema development, which Spillane et al (2002) define as: “Knowledge structures that link together related concepts used to make sense of the world and to make predictions” (p.394). A proactive strategy for managing tacit barriers is to influence individual and organisational schemas as well as to nurture links between the two (Butler, 1996; Spillane et al., 2002; Timperley & Robinson, 2002). Along this line of thinking, horizontal supports were mechanisms for aligning individual, school and cross-
school schemas. That sort of alignment has been referred to in the literature as cognitive alignment (Spillane et al., 2002).

In summary, common barriers to practitioner learning tended to be pro-actively managed mainly through vertical support mechanisms although there were some horizontal supports that complimented the vertical ones. A combination of strong design theories and sufficient attention to the adequacy of resources and support were features of vertical management strategies for overt barriers. There was a noticeable absence in the literature of references to overt barriers becoming problematic to practitioners, which suggests they were managed towards the top of the systemic hierarchy. In terms of horizontal supports for tacit barrier management, the placement of co-ordinators within and around schools was a common strategy. They appeared to help initiate some practitioner communities of practice, although the reason for their placement was primarily to support the vertical learning processes.

5.3. The Comparative Framework

The comparative framework presented in Figure 5 is made up of three basic elements: a set of connected learning contexts with each context having a priority learning focus; vertical and horizontal learning dimensions across the contexts; and, pro-active management strategies to overcome barriers to practitioner learning. Each of the three elements is explained in turn. Built into the explanations is the way that the elements will be used to create the learning process models in Part III.

5.3.1. Learning contexts

There are four learning contexts in Figure 5; a national policy context, a school improvement initiative context, a school context and a classroom context. They are represented by the set of concentric circles. The four contexts displayed are only examples. Using concentric circles made it possible to add as many contexts as is necessary for a particular schooling improvement initiative. Additionally, the boundaries of each circle not only differentiate one context from another, they also helped to distinguish different learning foci. In the framework, for instance, the school as a learning context is linked to organisational learning and the classroom context is linked to learning academic content knowledge and pedagogical practices.
Chapter Five: A Comparative Framework

Analysing contexts and learning foci may seem trivial when there are so many complex issues that need to be attended to urgently within the school improvement movement. However, analysing simple everyday things should not be overlooked in theory development about how and why practitioners are expected to learn the way they are.

![Diagram of Vertical Learning Dimension](image)

**Figure 5.** A Comparative Framework to Construct Learning Process Models.

5.3.2. **Vertical and horizontal learning dimensions**

The vertical learning dimension represented by the vertical arrow in the middle of Figure 5 refers to hierarchical connections and transfers of knowledge and skills between policy developers, researchers and practitioners. The downward thrust of the arrow relates to what practitioners are told to do by others higher up the hierarchy. The upward thrust of the arrow is about the feedback loops back up through the
hierarchy. The width of the arrow can be altered to represent what is happening in different initiatives. For instance, it can be widened to represent significant activity through the hierarchy or made thin if very little is happening.

The vertical learning dimension shows the nature of interactions and knowledge transfers through the hierarchy. These interactions involve deliberate learning through formal structures and they bring into play positional authority. The vertical dimension is particularly useful for understanding the role of government and its agents. They are necessarily positioned at the top of the hierarchy as they make education law, which impacts on every level of the system. However, the learning process models presented in the next chapter will reveal that the top end of the hierarchy is not the only place that government influence can be found.

The horizontal learning dimension represented by the horizontal line running through the bottom of the comparative framework in Figure 5 refers to non-hierarchical learning opportunities through connections with colleagues, friends and information nodes that are void of power and control issues. It accounts for much of the non-deliberate learning through informal structures, such as communities of practice. Authority is vested in those that have relevant knowledge rather than in the holder of particular positions. As with the vertical arrow, the width of this arrow is intended to be flexible so as to represent what is happening in any particular initiative. The limited influence of positional power and control in the horizontal dimension does not restrict government influence within it. The analysis of the SEMO project in Part III of the SEMO project, for instance, will reveal how government agents have learned to influence non-hierarchically through horizontal learning networks.

Finally, the interrelationship between the vertical and horizontal learning dimensions is represented in the framework in Figure 5 as a dotted line stretching diagonally from the vertical arrow to the horizontal arrow. This dotted line ensures that the analysis of the initiatives checks for interrelationships between the two dimensions. Are the two dimensions independent of one another or are they interdependent? As in the other elements of the framework, the dotted line can take many forms or not appear at all depending on what is happening. Analysing the interrelationships between the two dimensions was prompted by a comment made by Michael Fullan at a conference session at the American Education Research Association conference in Montreal in April 2005 (M. Fullan, personal communication, April 12, 2005). I presented an early iteration of the framework and
the models at the session. In the interactive discussion at the end of the presentations, Fullan claimed that understanding the interactions between vertical and horizontal learning levers and getting the right mix was one of the next big challenges facing those involved in trying to advance the school improvement movement.

5.3.3. Managing barriers to learning

The jagged line going from the outer circle into the centre circle represents barrier management in the comparative framework in Figure 5. Its jagged appearance is intended to represent interference across the learning system. The line of interference can be configured in many different ways depending on the presenting barriers and management strategies to deal with them. For instance, the example of barrier management in Figure 5 indicates that the problem of using practices that do not work is right across the system. It may be that in some initiatives the priority barriers are pro-actively managed in such a way that the line of interference is not disrupting what happens in classrooms, which means the jagged line would stop short of the classroom context. Alternatively, the line of interference may only be a problem within schools and classrooms and so the jagged line would be contained within those two contexts. Numerous barriers can be identified and represented in the models but the intention is to identify priority barriers that pose considerable challenges to learning effective schooling improvement. Those will undoubtedly vary across the three countries because of the different education systems and macro policy contexts.

5.4. Summary and conclusions

An iterative analysis process of sifting through a great deal of descriptive and evaluative information about the five international schooling improvement initiatives was used to find broad development and implementation characteristics relevant to learning effective reform practices. The analysis identified a set of common conditions about what and how practitioners were expected to learn the preferred reform practices. A key "what" condition revealed was standardised design theories and implementation practices. Three "how" conditions discovered were: getting practitioner buy-in to the theories and strategies, vertical and horizontal learning
opportunities and constantly checking the impact of the interventions on the students' achievements.

The set of common conditions was then transformed into the comparative framework, which had three vital parts: connected learning contexts with specified learning foci; vertical and horizontal learning dimensions; and pro-active strategies to manage barriers to learning. The thinking behind the three elements is that there is a background and a foreground to practitioner learning. The background is the contexts within which learning can occur through deliberate, spontaneous or chance situations. The foreground is the ways that learning is encouraged and the barriers that tend to impede learning. Those two foreground issues bring into play the vertical and horizontal learning dimensions and the management strategies to overcome barriers. Background-foreground thinking resulted in the vertical and horizontal learning dimensions and the barriers to be laid over the learning contexts in this framework.

A limitation in the design of the comparative framework is that it did not draw on conditions related to successful initiatives in New Zealand, which could arguably have revealed additional or conflicting features. Its features are based on what is happening in a few initiatives in the United States and England with strong evidence of effectiveness. The point raises issues about the relevance of the framework to the New Zealand context. This may seem particularly contentious given that the United States and English contexts are heavily influenced by policy mandates and the New Zealand context is not. That limitation is considered a minor one rather than a serious flaw. The main reason is because a critical design principle for the comparative framework was that its features had to be linked directly to characteristics attached to strong evidence of effectiveness. The five international initiatives used to inform it had such evidence (3.1. p.62) and there were no initiatives found in New Zealand with equivalent evidence (4.1, p.84). Had there been any, their characteristics would have been included in the design process. I was not prepared to compromise my evidence-informed design principle in this instance, hence the exclusion of the New Zealand-based initiatives. That decision did not mean I was going to use the framework to judge characteristics of New Zealand initiatives unfavourably. To the contrary, the framework provided a means by which I could find and explain interesting similarities and differences that are worthy of consideration to further develop New Zealand’s schooling improvement approach.
Chapter Five: A Comparative Framework

To conclude, the comparative framework developed in this chapter was constructed as an essential research tool to advance the theory development in this thesis. Through the process of developing the tool it became apparent that understanding the generation, transfer and use of knowledge through school improvement learning systems is essential to accelerating the learning of effective reform practices. Knowledge and skill generation, transfer and use, therefore, become a central thread in the theory development from this point on. In the next chapter (Chapter 6), for instance, the three learning process models reveal a sharp contrast between the way the knowledge transformation process occurs in New Zealand and in England and the United States.
CHAPTER 6
LEARNING MODELS

This chapter presents and explains learning process models for the seven school improvement initiatives selected for analysis. There are three models in total. The first one represents learning processes in England’s literacy and numeracy strategies (Figure 9), the second represents the processes in the four initiatives operating in the United States (Figure 10) and the third one shows those processes used in the two initiatives in New Zealand (Figure 11). It is important that the models are seen for what they are, namely, representations of learning systems in seven schooling improvement initiatives operating in three developed countries. They are not an attempt to synthesise the learning processes used across all school improvement initiatives in all developed countries. The chapter begins with the presentation of the three models and explanations of their component parts. Following those explanations are the findings from an analysis of the models to ascertain the sorts of learning connections that were being used within the initiatives to transform useful theoretical and practical ideas into effective reform practices.

6.1. The Learning Process Model for England’s Literacy and Numeracy Strategies

The three components of the learning process model for England’s literacy and numeracy strategies presented in Figure 6 are: (i) five learning contexts; (ii) a strong vertical learning dimension and only a few horizontal learning opportunities and, (iii) barrier management focused on easing the load on practitioners so they could focus on teaching and learning. There were no noteworthy connections between the vertical and horizontal dimensions so there is no diagonal line connecting those two connections.
6.1.1. Learning contexts

Five learning contexts were found to be transferring and spreading the preferred practices in England’s Literacy and Numeracy strategies. From the outside in, the five contexts are labelled in bold as the national policy context, the national literacy and numeracy centres of expertise, the local education authorities and consultants, schools and classrooms. The national centres served as centres of expertise and local authorities and consultants acted as bridging agents between the directors in those centres and the principals and teachers in schools. Things had to happen quickly after the elections because of the incoming government’s promise to urgently address the neglect of educational opportunity for disadvantaged students (Barber, 1998; Earl et al., 2000). Because of government’s urgency, policy developers capitalised on findings from the previous governments’ literacy and numeracy taskforces and trials to formulate the policy (Brown, 1998). That meant most of the policy learning had
been accomplished earlier, although one addition was to ensure that the national achievement targets and progress towards them were publicised much more publicly than had been the case in the past.

Learning foci in the other four contexts concentrated on transforming the policy mandates into practice. The learning focus in the national centres was about how to design a national intervention and monitor it to make sure the ambitious reform effort unfolded the way it was intended, i.e. "The core of the Strategies has always been the focus on the teaching of literacy and mathematics in schools" (Earl et al., 2003, p. 39). In turn, local education authority officials and consultants needed to learn how to support practitioners to absorb the pressure and bring the practices alive in the schools and classrooms. So the principals were constrained to learning about school-wide systems to support the application of the standard practices. Teachers were constrained to learning relevant curriculum content and pedagogy by the strict boundaries set down for the professional development workshop sessions (Department of Education and Skills, 2002).

6.1.2. Vertical and horizontal learning

A strong vertical learning dimension was the primary means of connecting the five contexts so that relevant knowledge could flow down and back up the hierarchy. There are three components to the vertical dimension. The first component is the strong downward flow of knowledge represented by the prominent vertical arrow in the centre of Figure 6. Its prominence represents the high levels of pressure and support that were applied in order to spread the reforms uniformly across the country (Earl et al., 2003). The second component of the vertical dimension is the feedback loops represented in Figure 6 by the thin arrow looping back up the hierarchy from one context to the next. The thinness of the arrow indicates that there was some feedback through the different levels of the learning system but it was not influential in terms of informing the direction of the reform practices or the targets (Earl et al., 2000). The developers were only open to greater flexibility once some good results were achieved.

The third component of the vertical dimension was the feedback loop to the policy context created by the international research team, which is represented by the arched arrow on the right of Figure 6. This feedback mechanism also did not challenge the
overall intent of the reforms. It was used to verify the achievement trends and to act as a critical friend to the policy developers and national directors to deepen their understandings about the nature of large scale reforms (Earl et al., 2000; Earl et al., 2003). As independent evaluators, they were able to look into the system and analyse the emerging policy and practice opportunities and risks. For instance, they validated Barber’s belief that large scale reform is useful to get some positive results on the board (Barber, 2000b).

Horizontal learning is represented in Figure 6 by two arrows running through the bottom of the model. One thin arrow spans the inner three contexts and another thicker one is contained within the classroom context. The longer thin arrow relates to the few formal horizontal supports set up to reinforce what was happening vertically, such as the principal forums co-ordinated by local authority managers (Earl et al., 2003). The thick short horizontal arrow sitting next to the long thinner one at the bottom of Figure 6 represents a particularly interesting aspect of the horizontal learning dimension in the English initiatives. That aspect is the informal networks of teachers that appeared to be operating behind the scenes. Brown et al (2000) claimed that an informal network of teachers appeared to spread enthusiasm for the reform practices before they were officially announced. Apparently, the spread happened without any formal structures or discipline and it happened swiftly. Had the swift spread of enthusiasm not happened, the level of dissent among teachers would have arguably been considerably higher than it was. The extent to which the reform designers intentionally capitalised on the teacher networks or struck it lucky is hard to tell. Regardless, knowing that such networks exist and tapping into them as a pro-active lever to accelerate things is intelligent reform thinking.

6.1.3. Barrier management

Efforts to sort out priority barriers appeared to centre on helping make the daily lives of practitioners more manageable. Although the unified bureaucracy applied considerable pressure on them, it also appeared to act as a pressure release valve for them. Pressure was associated with principals and teachers implementing the standardised practices effectively to achieve the national targets. The pressure release valve was related to distributing leadership for planning and evaluation across the learning system. For instance, strategic frameworks, lesson plan templates and
aggregated achievement information were organised by the national centres and support was given to practitioners to learn how to use them by Local Education Authority officials (Department of Education and Skills, 2003). The research team tracking the developments saw this arrangement as a critical component of effective strategic leadership: “Results of this study suggest that to be strategic, laterally distributed leadership must be embedded within a complementary vertical leadership structure. Given such a structure, those at the top need to focus on direction-setting practices and the provision of resources to make the implementation of large-scale reform possible. This was the main focus of the Prime Minister, DfES [Department for Education and Skills] and SEU [Standards and Effectiveness Unit] in the case of NLS [National Literacy Strategy] and NNS [National Numeracy Strategy], for example” (Leithwood et al., 2004, p.76).

A positive consequence of distributing leadership responsibility for planning and evaluation across the unified bureaucracy was that it alleviated the possibility of school level mismanagement. Practitioners did not have to sift through large amounts of information and create things from scratch. The international research team noted this particular point in terms of relief among practitioners that they were not being over burdened: “Head teachers and teachers often expressed relief that they had been given the NLS and NNS frameworks and curriculum materials to better cope with the pressure from national tests, Ofsted inspections, imposed targets and high workloads” (Earl et al., 2003, p.130).

6.1.4. Summary of the model for England’s Literacy and Numeracy Strategies

To summarise, the main learning mechanism for practitioners used in England’s Literacy and Numeracy initiatives was the vertical transfer of standardised practices from the national centres of expertise to the classrooms. The transfer saw practitioners relating to local education authority officials and consultants. Relationships appeared to be procedural to ensure that the practitioners understood the standardised practices and used them effectively in the schools and classrooms. There were a few horizontal learning opportunities developed for the practitioners and they were controlled by the local education authorities or consultants to resolve implementation problems as they arose. Barrier management was all about making things more manageable for practitioners so they would not be distracted from
implementing the standard practices. Practitioners, therefore, were treated as implementation agents of a pre-determined script to get all Year 11 students performing at expected levels in literacy and numeracy.

6.2. The Learning Process Model for the Four Initiatives in the United States

The configuration of the learning process model in Figure 7 for the initiatives operating in the United States is (i) a diverse set of learning contexts stemming from a prominent national policy context; (ii) a strong vertical learning dimension and a few horizontal learning opportunities in three of the four initiatives, the exception being the School Development Project; and (iii) barrier management centred on the same issue as that in England of helping practitioners focus more sharply on teaching and learning. As was the case with the English model, there is no diagonal line connecting the vertical and horizontal learning connections because vertical mandates dominated most, if not all, of the learning opportunities.

6.2.1. Learning contexts

The policy context at the top of Figure 7 has a strong presence because of the challenging national standards set down in the No Child Left Behind Legislation (United States Government, 2002) (details in 2.2). Learning contexts below the policy context are self-contained within each of the four initiatives. There are three different configurations of learning contexts represented. The one on the left of the model (representing the School Development Project) and the one in the middle (representing Direct Instruction and Success For All) closely resemble the English configuration in that there is a centre of expertise with co-ordinators connecting the centres to the schools and classrooms.

The set of contexts on the right of the model, which represent the learning processes set up in New York District #2, differ from the norm on two counts. The first difference is that the District Office did not represent a centre of expertise. Instead there was collaboration between the district office officials and the research team helping them move things forward (Elmore & Burney, 1997, 1998, 1999, 2000). For instance, the research team assessed the extent to which the district office
officials had instituted a process of continuous improvement and offered suggestions for further development upon Alvarado's departure (Elmore & Burney, 1998).

The second difference is that there was no bridging type context between the district office and the schools. The district office officials retained a direct relationship with the schools in the district by negotiating the schools' improvement plans with principals and by monitoring the practitioners' implementation of those plans. Consultants also worked directly with schools. Their involvement appeared to widen the district office team to create "a coherent, well-worked-out strategy for instructional improvement, focused mainly on, but not limited to literacy and
mathematics. This strategy emphasizes professional development for teachers and administrators, focused largely on high quality instructional practice (where “quality” is determined by explicit benchmarks and standards)” (Elmore & Burney, 2000, p.3). So the schools could not avoid focusing on instructional improvement. They were literally surrounded by officials and consultants with intervention expertise to make that happen.

6.2.2. Vertical and horizontal learning

The vertical learning dimension in Figure 7 is more prominent than the horizontal dimension, as was the case in the English-based model. In this case, however, there is no continuous arrow from the policy context to the classroom. It is disconnected between the policy context and the initiatives. That is because the No Child Left Behind legislation sets challenging standards but does not mandate the actual programmes that schools are expected to use (United States Government, 2002). That legislation and earlier versions of the United States’ legal framework governing Title I funding meant the task of doing schooling improvement has been devolved to the state, district and school levels of operation. The four initiatives represent the different ways that the devolution has been played out. Despite considerable opportunity for diversity, the developers of all four initiatives (S. Engelmann, Slavin and Madden, Comer & Alvarado) preferred to use the vertical learning dimension over the horizontal dimension to introduce reform practices and to get some loyalty among practitioners for using them.

There were three variations in the way that the developers used the vertical learning dimension. One variation was to get a standardised script into the classrooms in much the same way that the English designers did. Direct Instruction and Success For All used the vertical dimension in that way. A second variation was to hold the principals and teachers to account for implementing negotiated improvement plans, which was the case in New York District #2. The third variation in The School Development Programme was to use it to introduce standardised teams and rules of engagement within those teams to address students’ development needs.

All three variations included feedback loops which are represented in each of the initiatives in Figure 7 by a thin arrow looping upwards from one learning context to the next. The feedback loops relayed implementation issues and achievement
information to the developers in order to refine the type of support that the coordinators/consultants were giving to the practitioners and to track the overall outcomes of the initiatives. One director at the National Institute for Direct Instruction summarised how the feedback loops typically worked:

There are two separate data streams. For schools implementing the model, we have regular student performance data used to evaluate the state of the Direct Instruction implementation. Our on-site consultants (an Implementation Manager and Project Director) examine the data and engage the school’s leadership team in weekly problem-solving sessions over the phone. The second is field test data of programs (instructional materials) used by S. Engelmann and co-authors to evaluate new programmes or revisions of existing programs. (K. Engelmann, personal communication, January 8, 2004).

So there was a dual purpose of feedback loops within the initiatives to ensure that things that mattered to practitioners were being dealt with and that the bigger strategic picture was also being monitored.

The wider research community provided an additional feedback loop to the initiative leaders and to the federal government officials involved in distributing Title I funding by conducting independent research into the initiatives. That independent aspect of the reforms appears to have a strong presence in the United States, hence the separate context at the bottom of the model in Figure 7. The number of independent studies referred to in Borman et al’s (2003) meta analysis of comprehensive school reforms indicates the extent of independent research activity. From the 108 studies investigating Direct Instruction (42 studies), Success For All (48 studies) and The School Development Programme (18 studies), for instance, Borman et al (2003) calculated that about a third to a half of them were independent third-party studies. The combination of independent third-party feedback loops alongside the feedback loops set up within the initiatives acknowledge the extensive efforts made in the United States reform movement to use strong empirical evidence to drive developments.

There were some horizontal learning mechanisms used in all four initiatives but they aimed more at supporting vertical mandates than encouraging self-determining collegial networks. Horizontal learning opportunities for principals in New York
Chapter Six: Learning Models

District #2 are a case in point. In a paper informing the development of principals in the district, the research team stated:

“Among the earliest professional development tactics the district used was the creation of “buddy” arrangements among principals, in which groups and formal mentoring arrangements have evolved into principals working on common problems... They are focused on specific topics, generated either by principal committees or by district administrators, that are related to the district’s overall instructional focus in literacy and mathematics” (Elmore & Burney, 2000, p.4).

This case indicates a disciplined approach to developing horizontal learning opportunities for practitioners. Other examples such as peer coaching arrangements in Direct Instruction and Success For All reinforced this change principle. Hence the relatively narrow horizontal arrows at the bottom of the learning contexts for Direct Instruction, Success for All and New York District #2 in Figure 7. The teaming arrangement within The School Development Project appeared at first glance to give practitioners more freedom to learn in a non-hierarchical manner. They sat alongside special staff, parents and school administrators (managers) to prioritise and address students’ development needs (The School Development Programme Website, 2004). However, they were constrained by the standard rules of engagement and by the schools’ instructional and environmental development plans (Comer & Haynes, 1999). This is, therefore, another example of a consistent development approach of using vertical mechanisms as levers to determine the nature of horizontal learning opportunities for practitioners. In other words, the teaming arrangement was simply a softer version of positional authority setting the reform learning agenda than the mandated agenda-setting approach used by the developers of the English literacy and numeracy strategies.

6.2.3. Barrier management

In terms of priority barriers, getting some clarity around reform responsibilities for practitioners appeared to be a priority in all four initiatives in the United States. There was a concerted effort to make sure that practitioners learn more about their core business of teaching and learning as was the case in England (Engelmann, 2003;
Harwell et al., 2000; Success For All, 2003; The School Development Programme Website, 2004). There was also a drive from the developers of the initiatives that the role of the school principal be a key change agent in the reform process. That drive appeared to be an attempt to get school leaders as focused on instructional improvement as teachers were expected to be. In the case of the three comprehensive school reforms, that meant capacity building among existing school principals. They had approximately three years to learn about and then help institutionalise the reform practices before intensive support was withdrawn. In contrast in New York District #2, the attention on school principals appeared to be about creating a new pool of more capable principals by removing those that did not measure up to the superintendent’s expectations. It was also about training potential principals: “The Aspiring Leaders Programme addresses an emerging need in the district’s overall instructional improvement strategy. It builds up the pool of people potentially available for principalship in the district and it does so in a way that is consistent with the district’s views on the nature of the principalship and its core instructional goals” (Elmore & Burney, 1998, p.25). It seemed, therefore, that attempts by the developers to clarify reform responsibilities were about getting both the leaders and the teachers into the same instructional improvement mindset.

6.2.4. Summary of the model for the four initiatives in the United States

To summarise the learning process model for the four initiatives in the United States, the vertical learning dimension was once again the main change mechanism used to help practitioners learn the reform practices. It became the preferred lever for change via experts external to schools who had the capacity to design and spread effective reform practices. However, practitioners were not expected to fit into one overall learning system. Rather, they fitted into one of four self-contained learning systems that generated standardised practices to help them get their students to reach the national standards. In contrast to the English reforms, the practitioners in the United States related to the research community rather than the policy community to learn the standardised practices. Barrier management was about distributing management responsibilities so that the school-wide systems and preferred teaching practices would be sustained once the external experts exited the schools after two or three years of support. Practitioners, therefore, were once again thought of by the
developers as implementation agents but they were also given responsibility to sustain things once they were up and running.

6.3. The Learning Process Model for the Two Initiatives in New Zealand

The three components of the learning process model for the Numeracy Development Project and the SEMO project in New Zealand presented in Figure 8 include (i) four learning contexts with collaborations featuring in the top and bottom contexts, (ii) a balanced use of the vertical and horizontal learning dimensions, and (iii) barrier management focused on moving from data-free to evidence informed decision making. In this case, there is a connection between the vertical and horizontal learning dimensions represented by the dotted diagonal line to the right of centre in Figure 8. An additional section is included in the explanation of the model to describe this connection.

Figure 8. Learning Process Model for the Numeracy Development Project and the SEMO project in New Zealand.
6.3.1. Learning contexts

There are four learning contexts in the New Zealand-based model in Figure 8. Viewing them from the top of the model to the bottom, there is a national policy context, a school context, a classroom context and a learning networks context. The national policy context is one in which policy developers have had to work out how to influence practitioners to learn effective practice without the mandates developed in England and the United States. As the description of New Zealand’s macro policy context in Chapter 2 outlined (2.3), successive governments have developed a policy platform that has avoided using directive powers in the ways that their counterparts in the United Kingdom and the United States have done. In the foreword of the latest research report on the Numeracy Development Project, Professor Derek Holton from the University of Otago captures general support in the education sector for New Zealand’s policy preference for avoiding such directive powers (Higgins et al., 2005). He states:

There are at least two ways that reforms can be undertaken in education. One way is to provide set lessons for the teacher and set work for the students. The second is empowering the teacher through professional development. In implementing the Numeracy Development Project (NDP), New Zealand took the second of these options” (p.1).

A consequence of the preference among successive governments and educationalists like Professor Holton to influence practitioners rather than direct them is that no contexts such as centres of expertise have formed in between the policy contexts and the school and classroom contexts. They could have formed but they did not most probably because the self-managing law (New Zealand Government, 1989) encouraged schools and teachers to handle things themselves. What did emerge in New Zealand in the absence of centralised steering was targeted guidance through two forms of collaboration in the vertical and horizontal learning dimensions. By ‘targeted’ guidance, I mean the sort of help that is useful for specific achievement problems. It was not the generic sort of guidance typically offered through the plethora of national guidelines and resources which assumed that school boards and practitioners can independently discern between relevant and irrelevant guidance (See
2.3, p.40-48 for details). An explanation of the two forms of collaboration is presented in the next section.

6.3.2. Vertical and horizontal learning

The first form of collaboration is one operating in the vertical learning dimension. It is represented by the oval at the top of the model in Figure 8 in the policy context. The oval represents the policy developers, researchers, resource developers and publishers involved in designing the Numeracy Development Project. It provided school boards and practitioners an opportunity to join a more systematic approach to improvement than has been the norm in the vertical learning dimension (Higgins & Parsons, 2005). By joining the Numeracy Development Project, school boards and practitioners were choosing to reduce their autonomy to pick and choose as they saw fit in the teaching of numeracy. In other words, if they joined up they did not have to be the sole designers of what they did in math lessons. The collaboration did not expect local boards and practitioners to work through the maze of information available to design, implement and self-review numeracy improvement strategies. It organised numeracy information in such a way that effective intervention options are more visible. Rather than leaving lesson planning, for example, open ended, teachers were guided by easy-to-read reference booklets (Ministry of Education, 2004a, 2004b, 2004c) and a comprehensive website (Ministry of Education, 2006h) to transform information from the number framework into three suggested lesson formats, i.e. knowledge lessons, strategy lessons and a mix of knowledge and strategy lessons.

The vertical collaboration did not completely remove teacher and school autonomy in the same way that the script-driven practices did overseas. Even though the information was highly organised, teachers retained some control over what happened in the classroom. For instance, they were charged with the responsibility of deciding when to deliver knowledge or strategy lessons or a combination of both depending on the students’ learning needs (Ministry of Education, 2004c). A great deal more is said in the next chapter (Chapter 7) about how the vertical collaboration attached to the Numeracy Development Project helped teachers become more discerning in the delivery of their lessons.

The second form of collaboration which emerged in the horizontal learning dimension was learning networks, which constitute the fourth learning context at the
bottom of the model in Figure 8. To recap on the definition of learning networks outlined in the introduction, they are connected groups of reformers working together what to do and what not to do to solve achievement problems among disadvantaged students. They are called “learning” networks because they are focused on learning how to solve achievement problems. They function as communities of practice, which have been shown to be successful in facilitating useful learning connections within and between firms (Wenger et al., 2004).

Colleagues with a common interest exchange ideas and information without interfering with the authority and control of directors and managers of each individual firm. That dynamic was demonstrated in the learning networks that emerged out of the SEMO project in that they did not challenge the positional authority of boards of trustees to oversee reform developments within individual schools.

The learning networks in New Zealand evolved through cottage industry reforms such as the SEMO project. The oval and overlapping circles depicting learning networks at the bottom of Figure 8 represent the most effective learning networks that emerged from the SEMO project. As they operated in the district of Mangere, they are referred to as the “Mangere learning network”. That particular network is sketched into the model in Figure 8 as two-tiered collaboration. One tier is the oval at the bottom of the model which represents a policy-research-practice collaboration made up of local officials (policy), researcher-developers (research) and a network of lead practitioners (practice). They collaborated with one another in a non-hierarchical fashion to work out what knowledge and skills would be spread among governors and practitioners in the schools that participated in the network. The collaboration has become known as the Mangere AUSAD Management Team (MAMT) (Ministry of Education, 2000). “AUSAD” is an acronym for the Analysis and Use of Student Achievement Data. It was the knowledge area on which the collaboration wanted to focus and develop.

In some reflective comments on the need in this regard, the chairperson of the management team, Karen Mose, outlined the state of affairs that things were in when the collaboration was getting started in 2000 and the sort of help that was needed to move forward in this knowledge area:
I realised our practices were all over the place. We needed a theory-informed development approach to align teaching, management and governance practices. Instead of a facilitator I realised we needed a teacher, an analyst, a problem solver, a research literate individual. These thoughts were then shared with the lead-assessment teachers and the management team agreed we needed someone to challenge our assumptions, develop our skills in using achievement information, expand our thinking and enable us to become evidence-based decision makers (Mose & Annan, 2002, p.6).

In the next chapter (Chapter Seven) a great deal more is said about the type of evidence-based practices that Karen and other participants in the Mangere learning network’s policy-research-practice collaboration developed.

The second tier of collaboration is the series of circles above the oval in Figure 8 which represent research-practice collaborations. There were several of those collaborations operating concurrently under MAMT’s umbrella. There were the two reading comprehension interventions into Years 1-3 students (Phillips et al., 2001; Timperley, 2002, 2004a; Timperley, Phillips et al., 2003) and Year 4-8 students (Lai et al., 2004; Lai et al., 2006; Lai et al., 2005; McNaughton et al., 2004). There were also two others focusing on the effectiveness of Pacific bilingual and special needs classes respectively (Ministry of Education, 2002b). It is the two collaborations that formed around the reading comprehension interventions that are of interest here because they contributed to the promising evidence base attached to the learning network Mangere.

Those research-practice collaborations had limited power and control issues. A pivotal improvement principle underpinning them was localised problem solving. Researchers co-constructed the interventions with the practitioners rather than telling them what to do and how to do it. Lai et al (2006) articulated the principle when they completed a replication study of what was happening in Mangere in the nearby district of Otara.

The evidence-based problem-solving process, which involved a professional learning community of teachers, researchers and policy-makers, enabled teachers to fine tune their teaching practices in specific areas identified by the process. In other words,
Chapter Six: Learning Models

teachers became inquirers of their own practice, using evidence from student achievement and observations of current instruction to address the identified teaching and learning challenges to raise achievement. (p.5)

The localised component of the principle immediately prompts thoughts of teachers using a Number 8 Wire approach to solving problems (see p.2 for details). However, that possibility was countered by researcher-developers with expertise in analysing and using achievement data helping practitioners to problem solve.

The two-tiered learning network in Mangere not only created effective learning processes for practitioners but it also created a rich information source for local officials to pass on to national policy developers. The two arrows linking the local officials to the national policy developers on the right side of Figure 8 represent that connection. One particularly useful strategy used to pass on useful context-specific information for national policy development were the vignettes of learning conversations that occurred at a regular monthly meeting set up to govern the Ministry’s influence on developments in Mangere and Otara (Ministry of Education, 1997b). Several senior policy developers as well as local officials and lead researchers involved in the SEMO project attended the meetings to discuss progress, debate strategies for difficult situations and tease out what to do with interesting findings31. Findings presented by the researchers from the first evaluation report (Timperley et al., 1999), for example, caused considerable discussion about the Ministry’s acceptance of add-on programmes as a ‘new’ solution. The researchers saw that solution as a way of maintaining old problems rather than introducing new and useful solutions (Timperley & Robinson, 2000). Although Ministry officials argued that investing in such solutions was a necessary step to help some self-managing schools come around to valuing external help (Annan, 1999), the research finding informed broader policy developments over the next few years, such as the national literacy and numeracy strategies (Ministry of Education, 2003b). Such local vignettes counter-balanced the massive amounts of readily available general knowledge that national policy developers typically rely on as evidence to formulate policy.

31 Minutes of the Ministry’s SEMO overview group meetings held from October 1997 through to December 2002 are archived in the Ministry’s project office Te Puna O Te Matauranga in Manukau City.
It is important to note that not all the learning networks set up within the cottage industry reforms operated in the same way as the Mangere network. Intervening into classrooms was not the highest priority when the learning networks started forming through the cottage industry initiatives in the latter half of the 1990's. Building relationships between groups with a vested interest in improving schooling in disadvantaged communities was what mattered most (Sinclair, 1999b). Key groups included the Ministry officials, the leaders of the schools in the disadvantaged communities and local leaders of those communities. Building relationships between those groups was found to be a complex and challenging improvement task (Annan, 1999; Robinson et al., 2000). In relation to the SEMO project, each group had different understandings of how schooling works and of the roles of the other groups. Additionally, each group had difficulty agreeing on what their groups' respective role was in relation to the students' schooling.

The investment into relationship building as a priority task reinforced the resolve of government to ensure that the long-term alienation of disadvantaged communities from schooling should cease (Ministry of Education, 1997b). However, it was not schooling improvement work as defined in this thesis. It was more of a consciousness-raising exercise about not alienating whole or parts of communities from schools. Most, if not all the cottage industry initiatives were part of the consciousness-raising exercise. The SEMO project, for instance, was announced as a three-way partnership between the Ministry, 37 schools in the districts of Mangere and Otara and the community members of those two districts. A great deal has been written about the way the Ministry officials as well as the school and community leaders, researchers and developers grappled with the complexity of the partnership (Annan, 1999; Annan et al., 2004; Robinson & Timperley, 2004; Robinson et al., 2000; Timperley et al., 1999). The manager of the monitoring and support division in the late 1990's claimed that the priority given to developing positive relationships within the SEMO project and other cottage industry initiatives set New Zealand's school improvement movement up to sustain its developments well into the future (Sinclair, 1999b).

A consequence of working through the complexities of the partnerships in the cottage industry projects was that a variety of different learning networks formed in Mangere and Otara and in other districts. Some continued to concentrate on defining
and building partnerships (McCaulley & Roddick, 2001). Others have moved on to become intensely interested in classroom interventions as a means of developing more effective educational partnerships between teachers and students (Ministry of Education, 2004e, 2005c, 2005d, 2005e, 2005f). The most effective learning networks in relation to the definition of effective school improvement in this thesis were those that chose the latter pathway. They managed to shift their interest from the partnership itself to the interrelationships between teaching, learning and achievement. In other words, schooling improvement is about linking teaching to student learning which, in turn, needs to be linked to achievement ratings.

Researchers tracking the SEMO project urged the networks in Mangere and Otara to concentrate on understanding the causal links between those three factors (Timperley, 2002). The network in Mangere responded immediately by setting up and continuing to develop the Year 1-3 reading comprehension intervention and introducing the Year 4-8 one (Ministry of Education, 2002b). The networks in Otara followed suit shortly afterwards by replicating the Year 4-8 intervention in Mangere (Ministry of Education, 2003c). Those relatively quick responses, compared with other cottage industry initiatives, were probably due to the learning connections that they had formed with the researcher-developers.

However, there were no external accountability levers directing them to take that pathway. So it was a matter of partners coming around to this way of thinking. Arguing that the coming-around process in Mangere and Otara was quick has to be put into context. It was quick in terms of the speed at which participants in cottage industry initiatives in New Zealand more generally are coming around to concentrating on interrelationships between teaching, learning and achievement. An analysis of the evidence of effectiveness attached to many of the 28 cottage industry initiatives found in New Zealand would suggest about a third of them are still coming around to that focus. The coming-around process in this regard is another example of the long game which was reported in Part II to be in use to develop programme evaluation for New Zealand’s schooling improvement initiatives (4.2.2.1, p. 90-95). One outcome of the use of the long game in developing and implementing schooling improvement initiatives is that the practitioners’ professional lives, particularly the lead practitioners (mostly principals and middle managers) were made much more complex than that to which they were accustomed prior to becoming involved in
school improvement. This development is the exact opposite of attempts in England and the United States to simplify the professional lives of practitioners.

6.3.3. Connections between initiatives in the vertical and horizontal learning dimensions

An interesting aspect of the New Zealand-based model that was not present in the other two models is some strong learning connections that were made between initiatives in the vertical and horizontal learning dimensions. Those connections are represented by the dotted diagonal arrow on the right of the model in Figure 8. Both the Numeracy Development Project and the learning networks attached to the SEMO project developed such connections. Cross-initiative learning connections in relation to the Numeracy Development Project involved its leaders connecting with the leaders of another cottage industry initiative in an urban district called Manurewa, which is close to the districts of Mangere and Otara. The leaders of both initiatives decided to bring together the most useful learning processes from their respective initiatives. This meant implementing the national professional development programme through the cottage industry initiative. So the vertically-driven Numeracy Development Project contributed the standardised teaching practices and resources (Ministry of Education, 2004a, 2004b, 2004c).

Meanwhile, the horizontally-driven cottage industry initiative contributed its local infrastructure of lead teachers to intensify the standard facilitation offered through the national project (Manurewa Enhancement Initiative, 2004). That meant the lead teachers were given additional time and training to provide formative supervision of the teachers as they implemented the new reform practices. By combining national and local learning processes, there was a good balance of consistent implementation and understanding the context. Recent achievement results indicate that students involved in the two projects in Manurewa are doing better in numeracy than students from similar numeracy project schools across the rest of the country (Young-Loveridge, 2005, in Higgins, 2005).

An example of an across-initiative learning connection stemming from the SEMO project involved a transfer of research findings from that project to a vertically-driven national literacy professional development programme. The transfer was made by researcher, Professor Helen Timperley. She helped develop a strong empirical
knowledge base about effective school improvement from her involvement in researching the SEMO project (Robinson et al., 2000; Timperley, 2004a; Timperley & Annan, 2004; Timperley, Phillips et al., 2003; Timperley & Robinson, 2000, 2002; Timperley et al., 1999). She also became adept at providing formative feedback to the participants of the learning networks as they tried to ratchet up their efforts. As Timperley’s contractual obligations for those services were ending, she and some of her colleagues were contracted by the Ministry to review the professional development programme attached to the national literacy strategy (Timperley, Parr, & Higginson, 2003). They found that those involved in the programme perceived that improvements were being made but they did not have sufficient evidence to prove that to be the case.

This evaluation attempted, at the Ministry’s request, to find evidence of improved student achievement as a result of the Literacy Leadership initiative. Two thirds of the schools were able to provide the evaluators with some student achievement data regarding their initiative. The data that were provided, however, did not enable an independent judgment about improved student achievement to be made. This does not mean there were no improvements; it is just that we were unable independently to verify them. Principals and literacy leaders reported that they believed their class-based projects were successful because they focused on teacher practice and beliefs. Further, many believed the projects had worked because they perceived that reading levels had improved or that instruction was more focused. Over half the teachers reported that they had learned something that resulted in changed teacher practices. Most others indicated that it had reinforced or affirmed what they already knew. (Timperley, Parr, & Higginson, 2001, p. 1)

Subsequently, Timperley helped revitalise the national professional programme. She became an advisor to the national design team. She also helped the facilitators refine their skills and knowledge of the same inquiry-based learning approach that the Mangere learning network was developing. That meant perceptions of achievement were not considered an acceptable measure of improvement. Early results posted on the Ministry’s Literacy and Numeracy website indicate that the transformation of the national project is making a significant impact on many students’ literacy levels (Ministry of Education, 2006i), although the results are yet to be scrutinised by the academic peer review process or by third-party evaluation.
Chapter Six: Learning Models

There was no documented evidence found of any policy decision to develop a balanced vertical-horizontal school improvement learning system in New Zealand. Instead, it evolved. Any possibility of vertical domination was counterbalanced by New Zealand's legal framework for education (New Zealand Government, 1989), which emphasizes local ownership and development. The balance of vertical and horizontal learning led to an interesting spread of design and intervention learning across the various contexts. Although the various participants (i.e. national policy developers, boards of trustees, practitioners, researcher-developers and local officials) had regular learning agendas to adhere to, they all became involved in school improvement design and development processes at their various levels. For instance, those national policy developers involved in the monthly governance meetings for the SEMO project were also involved in other priority national policy development processes. While at the micro-level of operation, lead practitioners involved in the learning networks attached to the SEMO project were helping to co-construct interventions at the same time as fulfilling their regular schooling duties.

6.3.4. Barrier management

A barrier to practitioners learning effective reform practices that the Numeracy Development Project and the SEMO project have helped to address is a general acceptance of data-free decision making in schools. By "general", I mean that the acceptance appeared to be widespread across the education system. In the vertical dimension for instance, the self-managing law (New Zealand Government, 1989) required policy developers to accept the way school boards chose to implement the national curriculum. In turn, the national assessment policy (Ministry of Education, 2001) compelled schools' boards to accept the way that their practitioner-employees selected and used the various assessment tools available to them. Additionally, several reviews of the SEMO project at the time when the most effective learning networks were forming in Mangere found that practitioners in the district tended to

---

32 The senior manager of the Ministry's team involved in SEMO, for instance, was also contributing heavily to policy developments in Special Education and Early Childhood. (Ministry of Education, 2003b).

33 Although job descriptions for lead practitioners involved in MAMT were altered to include the schooling improvement work (Ministry of Education, 2003a), they were not excused regular duties, such as curriculum development, budget responsibilities, human resource management and school-community relationships.
accept one another’s ways of operating regardless of their effectiveness (Annan et al., 2004; Mose & Annan, 2002; Timperley, 2002).

An example of data-free decision making is reported in a Ministry report documenting the inputs, outputs and outcomes of the SEMO project after five years of implementation (Annan et al., 2004). There is a case study in the report of one school referred to as School 14 that was found by local officials to be underperforming for most of those first five years. Yet the board and senior managers refuted the claims without any data to substantiate their stance. The review states,

School 14 is one extreme example of a school resisting change, and it took four and a half years before positive results emerged. The principal and senior managers consistently claimed high performance and the board supported the claims of their professional leaders, even though available data indicated inadequate school performance. [Officials attached to] SEMO consistently claimed that the school performance was seriously inadequate. Over four years, Ministry support for the school’s solutions incrementally decreased in favour of serious governance and management support which was recommended by SEMO [officials] and its contracted consultants. The Minister eventually dissolved the board and in January 2000 appointed a commissioner under Section 78N of the Education Act, to address the performance issues. (Annan et al., 2004, p.38)

Not only did the board and senior managers of school 14 refute the Ministry’s claims without evidence to support their stance, but those practitioners in the school that may have agreed that there were serious problems remained silent throughout the five-year intervention. Those actions coupled with a silence about the situation from surrounding schools’ boards and professional colleagues added weight to School 14s’ leader’s complacency about their performance.

A consequence of the systemic acceptance was that school boards and practitioners could cast aside new and more effective ways of operating and return to personally-valued but less effective practices (Annan et al., 2004; Timperley, 2002). Underachievement, therefore, remained acceptable even though there was sufficient evidence to persuade them otherwise. One source of the general acceptance was the limited external accountability levers available at that time to effect change. Those limitations stemmed from the high levels of autonomy apportioned to school boards
and practitioners to administer their own affairs (New Zealand Government, 1989). Feedback from senior policy developers claimed that the high levels of autonomy were intended to liberate school boards and practitioners from the all-encompassing decision-making powers of the old Department of Education (National manager with oversight for cottage industry initiatives, New Zealand Policy Developers' Feedback, 2005). They were indeed liberated. They developed a strong sense of ownership over their schools as well as a sense of freedom to deliver and contract services in line with business practices (Butterworth & Butterworth, 1998; Wylie, 1999). However, an unintended consequence of the administrative liberation was complacency within schools that they had the knowledge and skills to deal with any situation that arose (National senior policy analyst associated with the SEMO project, New Zealand Policy Developers' Feedback, 2005). As it happened, the complacency was misplaced within those schools struggling with the complex and difficult task of solving the underachievement problem among disadvantaged students.

6.3.5. Summary of the model for the two initiatives in New Zealand

To summarise, the learning process model for the two initiatives in New Zealand revealed a more balanced use of the vertical and horizontal learning dimensions than was the case in the international reforms. A feature of the preference was an attempt to develop expertise in designing and developing school improvement initiatives throughout the education system rather than housing it in a centre. Local knowledge development via the cottage industry initiatives was given as much value as the importation of knowledge via the national professional strategies. That feature is in sharp contrast to the international initiatives which preferred to import knowledge via external experts. Attempts to distribute expertise in developing schooling improvement initiatives across New Zealand's education system have made lead practitioners' lives far more complex than their international counterparts. They are now expected to become involved in designing and developing sophisticated improvement interventions as well as completing regular leadership duties within the schools.
Chapter Six: Learning Models

6.4. Learning Connections

The vertical arrows in the three learning process models were mainly about connections that helped develop knowledge for practitioners and the horizontal ones were about connections that helped develop knowledge with practitioners. As the information in the previous sections of this chapter only alluded to the different sorts of learning connections, a deeper investigation of them was warranted to better understand the knowledge transformation process.

6.4.1. Learning Connections that Managed Knowledge for Practitioners

Learning connections that managed knowledge for practitioners occurred within the vertical learning dimension in all three learning process models. Hence, they are referred to as ‘vertical learning connections’. They were an important part of a process of importing explicit knowledge through the vertical learning dimension in five of the seven initiatives: England’s Literacy and Numeracy strategies (in Figure 6); Direct Instruction, Success For All and New York District #2 in the United States (in Figure 7) and the Numeracy Development Project in New Zealand (in Figure 8).

Explicit knowledge is that which has been codified and communicated verbally or in writing (Lam, 1998; Wenger et al., 2004). It is the sort of knowledge that can be aggregated and spread widely among people other than those who created it. It can also be easily stored and referred to whenever necessary. The importation process in those five initiatives was about inserting relevant knowledge and effective practices into the minds and actions of the practitioners and checking that they were using them as intended. It was also about discrediting existing thinking and actions that were in conflict with the reform ways of doing things. The five initiatives fell into two groups in terms of the nature of the vertical learning connections.

One group, made up of England’s literacy and numeracy strategies, Direct Instruction and Success for All, adopted learning connections consistent with those used in a machine bureaucracy. The term ‘machine bureaucracy’, which comes from organisational theory (Lam, 1998), is characterised by standardisation, clearly defined divisions of labour and close supervision. It is the type of bureaucracy often used in mass production firms to control the production of standardised goods. What was happening in England’s strategies, and in Direct Instruction and Success For All
mirrored the characteristics of a machine bureaucracy in that these initiatives nurtured a scripted importation process. Figure 9 shows the points at which critical vertical learning connections occurred within the three initiatives to complete the scripting process. Officials and their agencies (in England) and expert researcher-developers (in Direct Instruction and Success for All) developed integrated knowledge banks at the top of the hierarchy and those operating at the bottom were only exposed to those relevant to their particular roles in the production line. Practitioners did not have to design anything or sort through lots of information. They just had to teach the way they were told from above and use the content and resources that were developed for them.

<table>
<thead>
<tr>
<th>Knowledge Generation</th>
<th>England’s Literacy and Numeracy Strategies</th>
<th>Direct Instruction and Success For All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Generation</td>
<td>National Directors &amp; senior officials</td>
<td>Researcher-developers in Centres of Expertise</td>
</tr>
<tr>
<td>Knowledge Transfer</td>
<td>LEA managers and consultants</td>
<td>Coordinators/Consultants</td>
</tr>
<tr>
<td>Knowledge Use</td>
<td>Principals, teachers and students</td>
<td>Principals, teachers and students</td>
</tr>
</tbody>
</table>

Figure 9. Learning Connections in the Three Initiatives Using a Machine Bureaucracy

A difference worth noting between the official-driven machine bureaucracy and those driven by researcher-developers is the intended time-span for using extremely constraining vertical learning connections. Blair’s labour government had no intention of setting things in stone for the medium to long term. Barber explained that a concentration on vertical connectivity was considered a phase in a continuous reform process, “The policy principle of intervention in inverse proportion to success is being applied steadily. If our overall strategy works, as more schools succeed they will have greater autonomy and reward. Ultimately each school would have very substantial autonomy” (Barber, 2000a, p.23). So a strong emphasis on vertical
learning connections was seen as a way of breaking away from past ineffective practices but it was not considered a sustainable solution. In contrast, once schools had chosen to participate in Direct Instruction or Success for All, the developers expected them to use the particular scripting process on offer as a permanent fixture. This implied that the developers of Direct Instruction and Success For All were confident that the knowledge banks in their centres of expertise were sufficiently well developed to sustain a long term solution.

The other group of initiatives using an importation process, namely New York District #2 and the Numeracy Development Project developed learning connections commonly used in a professional bureaucracy. A 'professional bureaucracy', another term that comes from organisational theory (Lam, 1998), is characterised by practitioners being trained to acquire standardised knowledge and skills and then applying it with a degree of autonomy. It is the type of bureaucracy associated with universities and hospitals. In relation to knowledge management within a schooling improvement initiative, it implies the existence of integrated knowledge banks at the top of the hierarchy and practitioners at the bottom accessing whichever banks they consider relevant to their specific teaching duties. Consequently, knowledge management for practitioners in New York District #2 and the Numeracy Development Project was not quite as constraining as it was in the initiatives using a machine bureaucracy. They did not have to adhere to one particular script but they were constrained by the boundaries of the principles, planning frameworks and resources developed for them.

Figure 10 reveals the nature of the vertical learning connections attached to the professional bureaucracies operating within the reforms in New York District #2 and in the Numeracy Development Project. They involved a much greater amount of collaboration than was the case in the initiatives that used a machine bureaucracy. Within New York District #2, for instance, the district office officials, the research team, principals and consultants collaborated at the top of the hierarchy to generate the schools' improvement plans. Similarly several different groups collaborated in the Numeracy Development Project to produce the number framework as well as the planning, implementation and evaluation resources for practitioners. Then at the bottom end of the hierarchy, similar high levels of collaboration helped practitioners make professional judgements about implementing the reform practices.
A particularly interesting finding in relation to the use of a professional bureaucracy was the inclusion of the students in the vertical learning connections in the Numeracy Development Project. Practitioners in that project not only had to acquire and apply knowledge prepared for them, but they also had to get in touch with the students’ thinking about solving number problems in order to plan relevant lessons (Higgins, Parsons, & Hyland, 2002; Ministry of Education, 2004a). That meant that teachers had to help students make some of their tacit knowledge explicit before they could use the reform practices effectively. Tacit knowledge is that which is implicit in the routines, conversations and experiences of groups of people operating in a given context (Lam, 1998; Wenger et al., 2004). Consequently, vertical learning connections in the numeracy project that were critical to a successful importation process did not stop at the teachers as was the case in the other initiatives. They were extended to include the relationship between the teacher and the student. That connection became just as critical for teacher learning as it was for student learning. This finding suggests that teaching practices which tap into the ideas, thinking and beliefs that reside in the minds of the students can be just as critical for teacher learning as it can be for student learning.

Figure 10. Learning Connections in the Two Initiatives Using a Professional Bureaucracy.
6.4.2. Learning connections that develop knowledge with practitioners

Learning connections that developed relevant knowledge with practitioners occurred within the horizontal dimension in relation to the School Development Programme in the model in Figure 7, p. 141) and the SEMO initiative in the Model in Figure 8 (p.146). Hence, they are referred to as ‘horizontal learning connections’. They are a non-hierarchical mechanism for developing knowledge locally with practitioners. The local knowledge development process is about making explicit practitioners’ tacit knowledge in order to work out what and how to operate more effectively. It is the same process as that used to expose the students’ mathematical thinking in the Numeracy Development Project only in this case it relates to the adult practitioners. It is a matter of drawing out what practitioners already think and know and altering or adding to their knowledge and skill sets depending on the achievement problems that they are trying to help solve. Because tacit knowledge varies from context to context, it is difficult to transform into a prescribed blanket solution (Lam, 1998). One has to interact with those that generate tacit knowledge in order to make it explicit and learn from it. Consequently, personal contact with practitioners is essential. Greater understanding in this area is particularly important for the New Zealand context because the cottage-industry style of school improvement invests heavily into localised collaborations.

Horizontal learning connections occurred in the School Development Project and the SEMO project through communities of practice. Figure 11 shows the nature of the horizontal learning connections that formed within The School Development Programme and the SEMO project. Those attached to the School Development Programme occurred within communities of practice developed within individual schools. They were between practitioners, specialist support staff and parents within the school planning and management team (Team 1), the student support team (Team 2) and the parent team (Team 3) to identify and address students’ developmental needs. It was a wrap-around support approach which positioned the teams as the principle change agent. Those sorts of collegial planning and support teams commonly sit alongside hierarchical teams within individual organisations to perform complementary knowledge management functions (Wenger et al., 2004). One of the key differences between collegial and hierarchical teams is that the collegial ones tend
to focus directly on knowledge and learning whereas hierarchical teams have to concern themselves with a broader range of tasks, including administration and line management responsibilities. So horizontal learning connections of the sort that were set up through the School Development Programme can help practitioners stay focused on teaching, learning and achievement.

Figure 11. Horizontal learning connections.

In contrast, the horizontal learning connections attached to the SEMO project stemmed from the strong connections formed between lead practitioners, researcher-developers and local officials involved in the learning networks. It is represented by the dark-shaded triangle drawn between those three groups in Figure 11. The octagon at the top of the triangle represents lead practitioners joining the network from several schools. Additional connections then branch out from the darkened triangle at two points. One point is at the top of the triangle where the lead practitioners developed horizontal learning connections with the teachers and students in the schools within which they worked. Those connections were in the form of formative supervision to
make sure that the knowledge and practices agreed to in the inner triangle were adopted and used back in the individual schools. The other point is at the bottom of the triangle where the local officials and research-developers created learning connections with national policy developers (see 6.3.2 for details).

An interesting finding in relation to the horizontal learning connections attached to the two different forms of communities of practice is that they nurtured collegial learning and accountability among practitioners and other groups involved in the initiatives. By collegial I mean that they learnt from one another and held each other to account without a power relationship giving one group the upper hand. They had time to try things out and come around to agreeing on the practices that they would commit to. It was an intrinsic change principle which fits comfortably with New Zealand’s policy commitment to a long game. That is not to say the collegial arrangements left practitioners to their own devices. Membership of the communities of practice ensured that practitioners were also learning from other groups deemed by the designers of the initiatives to be important in working out what and how to do things.

6.4.3. Summary of vertical and horizontal learning connections

An analysis of the learning process models revealed two types of learning connections within three different sorts of infrastructures operating within the seven initiatives. Vertical learning connections were found to be a feature of the initiatives using a machine bureaucracy (England’s Literacy and Numeracy strategies, Direct Instruction and Success For All) and those using professional bureaucracies (New York District #2 and the Numeracy Development Project). In contrast, horizontal learning connections were used to develop the communities of practice in the School Development Programme and the SEMO Project.

6.5. Summary and Conclusions

Findings associated with the learning process models presented in this chapter provided some useful information to answer the third research question about development and implementation characteristics relevant to accelerating the learning of effective reform practices. As it happened, there was a great deal of information so
I presented it in two parts; a general analysis which explained the learning models and a more focused analysis outlining the learning connections used within the seven selected initiatives. The explanations of the models revealed a variety of processes. There was a distinct difference between the processes preferred overseas and those preferred in New Zealand. The international initiatives were found to prefer vertical learning mechanisms whereas the two New Zealand initiatives represented a preference for a balance of vertical and horizontal learning mechanisms. Additionally, the international initiatives tried to make practitioners' lives more manageable whereas cottage industry initiatives like the SEMO Project made things much more complex for them in New Zealand. These findings reiterate the point made in Chapter Two that schooling improvement developments tend to be contextualised within the politics and education of each country.

The explanation of the model representing the New Zealand initiatives also revealed some interesting connections between initiatives operating in the vertical and horizontal learning dimensions. The two examples cited show that across-initiative learning connections can happen in different ways. Learning connections between the Numeracy Development Project and the cottage industry initiative in the district of Manurewa were an example of design leaders from both initiatives combining national knowledge banks with local infrastructure. Whereas the transfer of knowledge from the SEMO project to the National Literacy Strategy exemplified how an individual researcher can be an influential connector for reform groups in need of new ideas. These sorts of developments across the learning dimensions are critical to the theory development in Part IV to help accelerate the learning of effective reform practices in New Zealand.

The more focused analysis of the learning connections found within the seven initiatives also found vertical or horizontal preferences. A heavy reliance on vertical learning connections within a machine bureaucracy implies a lack of confidence in practitioners' ability to innovate. It could be argued that those preferences reduce the art of teaching from a professional activity to a craft practice. Although that sort of argument stands up generally for factory-model initiatives in education (Kliebard, 2002), I support the view of some researchers that there is a need for an importation process of some sort in certain circumstances (Rowan & Correnti, 2006). The initiatives served to get something positive going across large numbers of schools where there had been little traction in the past in England and the United States.
Additionally, they got significant gains in student achievement. So vertical learning mechanisms within a machine bureaucracy which help import explicit knowledge into the minds and actions of practitioners do have their place, particularly when there is a need to get widespread action relatively quickly in an area that appears to have been stuck for some time.

Practitioners engaging in vertical learning connections attached to professional bureaucracies and those using horizontal learning connections within communities of practice had more control over what they thought and did to make improvements than their counterparts in the machine bureaucracies. In the communities of practice attached to the School Development Programme and the SEMO project, for instance, no one group directed the rest. It was the participating groups' understandings, experience and willingness to learn from one another that made them valuable members of the change process rather than their privileged position in a hierarchy or in academia. However, distributed power and control did not engender loose arrangements. To the contrary, the communities of practice were highly disciplined arrangements that led to standardised ways of doing things. This end-point was identical to the vertical learning environment but the way that the horizontal learners got there was far more self-determined.

The use of horizontal learning connections as an improvement mechanism assumes that practitioners, particularly lead practitioners such as middle managers, have the time and wherewithal to become involved in sorting through vast amounts of information, designing schooling improvement initiatives and implementing them. Expectations of lead practitioners in the learning networks associated with the SEMO project, in particular, were exponentially higher than those in any other initiative. The high expectations were set partly by the Ministry officials and researcher developers agreeing that the middle managers should take a lead role in developing the learning networks. They were also set by the middle managers accepting the lead role and then constantly raising the bar in what they were trying to accomplish.

To conclude, the different models of learning processes and different forms of vertical and horizontal learning connections suggest that there are several different ways of helping practitioners learn effective reform practices depending on the circumstances. For instance, in the English circumstances where practitioners had limited capability to move forward it appeared best to build capacity through a directive linear fashion, i.e. the use of a machine bureaucracy and to consider
horizontal learning networks once some positive results had been registered. In contrast in New Zealand, where the legal framework limited systemic direction, it appeared best to operate a professional bureaucracy and learning networks concurrently. Although these findings once again reiterate the contextual nature of schooling improvement, there is the additional point made in this chapter that New Zealand is open to both vertical and horizontal learning mechanisms and connections. So it seems sensible for New Zealand's reformers to consider using those learning processes and connections that other reform groups have used overseas to solve some of the more resilient achievement problems, such as underachievement among many students of Maori and Pacific Island origin.
This chapter reveals the findings of a deeper analysis of the learning connections between the officials, researcher-developers and practitioners collaborating with one another to develop the Numeracy Development Project and the Mangere learning network. The collaborations are a New Zealand-specific phenomenon within the seven initiatives analysed in this thesis. I believe they have developed in New Zealand because the officials participating in them have not had rule-governed authority typically experienced by their counterparts in England and the United States. A main reason for this situation is New Zealand's policy avoidance of large external accountability systems which was discussed in Parts I and II. Whilst that avoidance has made it difficult to produce high quality evidence of the effectiveness of schooling improvement initiatives in New Zealand, it has allowed some interesting learning connections to develop ahead of accountability connections between officials, researcher-developers and practitioners.

At the moment, what happens in the three-way collaborations is implicitly known by those involved in them. There is little written about them, probably because the researchers involved in them who are most likely to write about such developments have been focused on publishing information about their dealings with the practitioners participating in the projects. That is a complex task in itself, which produces sophisticated reports such as the Picking up the Pace report (Phillips et al., 2001), let alone talking about the involvement of officials in the learning process. An additional reason for the lack of explicit information about these particular collaborations is the Ministry’s preference to stay in the shadows and build leadership among the research and development and practice communities. That preference has meant there has been little investment into producing publicly available documents which explain how officials have connected with researchers, developers and practitioners to ensure effective reform practices are learned and implemented. This chapter hopefully begins to fill the void by making explicit two collaborations which I believe represent powerful design mechanisms for developing effective schooling improvement interventions.
7.1. A Vertical Policy-Research-Practice Collaboration

The learning connections between the officials, researcher-developers and practitioners in the Numeracy Development Project represent a vertically-driven policy-research-practice collaboration within New Zealand's school improvement movement. The nature of those connections and the outcomes from them are presented in the diagram in Figure 12. The diagram is best read from bottom to top through the trajectories of the three groups listed at the bottom of Figure 12. The main participants fell into three main groups; the facilitators on the left, the collaboration of policy developers, researchers, resource developers and publishers in the middle and the teachers on the right. Information about the three groups and about the nature of their interactions with one another is organised into three phases. In boxes at the bottom of the diagram is a description of their pre-intervention development needs from the past. In the middle of the diagram there are a series of ovals which show the nature of the learning connections among various groups. The three outer ovals represent the three groups and the set of reform practices that they used in the learning process. The two inner ovals represent the third and fourth space communities-of-practice and what was happening within them to achieve the post-intervention understandings in the three future-focused boxes at the top of Figure 12.

7.1.1. Pre-intervention development needs

A key researcher contracted to inform the development and implementation of the project claimed that all three groups came into the project with specific development needs (Higgins, 2001). Facilitators tended to bring with them a tradition of putting teachers through generic professional development programmes at particular stages of their careers. They needed to shift the locus of control from themselves to the teachers so that the teachers developed an interest in understanding theories and practices in teaching number (Hughes & Peterson, 2003). The argument was that because most teachers have not been interested in theory-practice relationships they have become stuck on written algorithms as the
Facilitators

Policy developers, idea researchers, resource developers & publishers

Past

Silo development of numeracy policies, resources, publications & research

Facilitators

Policy developers, researchers, resource developers & publishers

Teachers

Strong content knowledge and pedagogy in number

Future

Strong content knowledge and pedagogy in number

Policy developers, researchers, resource developers and publishers co-constructed the design, development and evaluation of the project

Third space

Facilitators and teachers
- modelled and observed each other conducting knowledge and strategy lessons
- talked, thought about and used resource booklets and other resources to match lesson content and pedagogy to students number knowledge and strategies

Fourth space

Policy developers, researchers, resource developers, publishers, facilitators and lead practitioners
- Agreed on domain knowledge: content knowledge and pedagogy in number (numeracy teaching model)
- Developed a number framework, diagnostic survey tools and lesson guidelines

Context-specific guidance of lesson content and pedagogy

Aligned development of numeracy policies, resources, publications & research

Delivered generic professional development programmes

English, Netherlands and Australian mathematics communities

Figure 12. The Policy-Research-Practice Collaboration in The Numeracy Development Project.
only way of teaching students how to solve number problems. A major problem with that situation was that it was restricting the students' thinking about appropriate solutions to solve mathematical problems. One teacher's reflections on the over-reliance on algorithms showed how students' understanding about number was being ignored at the expense of efficient methods: "If they [the students] are going to do it the algorithm way, a lot of what we teach them is just a method and so they do not necessarily have a great understanding" (Higgins, 2001, p. 39). Alongside those two groups, the policy, research, resource and publication communities tended to work in silos (Higgins et al., 2002). Those developmental needs made it somewhat inevitable that the project was going to evolve as the three groups learnt from one another.

7.1.2. The fourth-space community of practice

The activities in the fourth and third spaces helped each of these groups work on their development needs in order to improve teaching and learning in number lessons. The fourth-space community of practice was where important overarching reform decisions were made. An analysis of the acknowledgments in several key publications found that the membership of this fourth-space community of practice was a complex arrangement (Ministry of Education, 2004a, 2004b, 2004c). Those participants from the professional communities listed in the oval at the bottom of the diagram in Figure 12 represented the core of the design and development team. They included representatives from the policy, research, resource and publication communities. They did not form a centre of expertise as was typically the case in the vertically-driven international initiatives. Rather, their connections represented collaboration among interested parties from New Zealand's mathematics community who had some expertise to contribute to the design and development of the project. Those parties also had a vested interest in seeing the project help improve New Zealand's mathematics ratings against other OECD countries. For instance, success mattered for the policy developers in terms of government making credible investments just as much as it mattered to the companies publishing the teaching
resources in terms of their market credibility with schools (Higgins & Parsons, 2005; Higgins et al., 2002).

The design and development team did not stop at collaboration between interested parties in New Zealand. That core team also developed learning connections with representatives of mathematics communities in England, the Netherlands and Australia to help them along the way (Ministry of Education, 2004a, 2004b, 2004c). These extended connections meant that a multitude of communities were contributing to the design and development of the project. Some participants stayed in one of those communities while others moved from one to another. It was as if some could add greater value by staying in one space whereas others were of more value moving from one space to another. For instance, Dr Joanna Higgins from Victoria University of Wellington College of Education stayed in the research community to critique and inform the publications, processes and outcomes (Higgins, 2001, 2002; Higgins et al., 2004; Higgins et al., 2005; Higgins & Parsons, 2005). At the same time, developers like Peter Hughes from The University of Auckland seemed to move among various activities. He made important content contributions to the number framework, the diagnostic interview tools and the resource booklets and produced important background information for the professional development programme and helped train the facilitators (Ministry of Education, 2004a, 2004b, 2004c).

Meanwhile two officials from the Ministry attached to the project worked to make sure that the various activities in the third and fourth space communities had a common and consistent policy foundation. Four foundation policy principles were outlined in a conference paper written by the lead researcher and one of the officials (Higgins & Parsons, 2005). The writers were reflecting on the way the leaders of the Numeracy Development Project have gone about transferring ownership of the project from the centre to practitioners. The first three principles focus on pedagogy and teacher learning while the fourth one centres on systemic connectivity. The principles are: (i) increasing the sophistication of mathematical ideas and teaching and learning strategies among teachers and students; (ii) ensuring that teaching decisions are informed by evidence of students’ thinking; (iii) constructing teacher knowledge in their own context of practice; and (iv) adhering to a “participatory dynamic that arises from the collective agency of a complex network of overlapping groups participating in the project” (Higgins & Parsons, 2005, p.6).
Chapter Seven: Policy-Research-Practice Collaborations in New Zealand

It is that latter principle which underpins the fluid movement and interactions among the participants in the fourth-space community of practice operating in the Numeracy Development Project. Higgins and Parsons (2005) expand on the nature of the network:

The key mechanism for the development of the collaborative process is a network of nationally coordinated groups that include a range of expertise drawn from the mathematics education community. Some of them have been established for specific purposes, that is, they are “fit for purpose”, and their brief has been to address specific policy formulation, implementation or evaluation/research issues involving the schemas and resources of the project. Other groups are ongoing and their primary focus is aspects of the evolving design of the project. For example, the development and refinement of the explanatory conceptual framework, the diagnostic interview and the teaching model have occurred over the course of the project, incorporating new research and feedback from teachers and facilitators. (Higgins & Parsons, 2005, p.8)

Further explanation of the network in those conference proceedings and an attached appendix explaining the make up and purpose of various groups (Higgins & Parsons, 2005, p.12 of Appendix A) did not indicate that fluid movement within the network was consciously planned so that some members operated across various communities and others stayed in one. Higgins and Parsons appeared to be writing about the phenomenon after the fact. The extent to which fluidity within networks such as this one can be planned rather than just letting learning connections evolve is worthy of further investigation. It may be that planning for fluidity constrains participants’ spontaneous urges to connect at the moment when they most need help to solve a presenting problem. On the other hand, it may make explicit the value of fluid movement within networks and encourage more learning connections to occur.

Among the numerous design and development decisions that the members of the fourth-space community of practice made, two were particularly important for the practitioners to learn effective reform practices. One was that they agreed that content knowledge and pedagogical knowledge in number was the priority domain knowledge that needed to be spread among practitioners (Higgins, 2001). The second was to
develop a framework, a set of standardised practices and a set of tools to help the learning process in the third space and, more importantly, to help practitioners work more effectively in their classrooms (Ministry of Education, 2004a, 2004b, 2004c). They were important decisions above others because they created a sharply-focused intervention as opposed to a general one. They also ensured that practitioners had supports to plan, implement and evaluate their instructional effort (National senior advisor for numeracy, New Zealand Policy Developers' Feedback, 2005). Those supports included the use of common assessment tools across the schools and the storage of aggregated achievement information (from the diagnostic surveys) in electronic databases in consistent ways for analysis and reporting purposes.

Consistency in assessment management across schools was a bold move made by the design and development team. They were encouraging schools to relinquish their own assessment tools in favour of common ones, which could have been interpreted as an attempt to set up unhelpful external accountabilities. Feedback from one policy developer attached to the project indicated that apart from a little dissent among some principals from intermediate schools about that agenda in relation to Year 8 and 9 results, that opposition did not happen (National senior advisor for numeracy, New Zealand Policy Developers' Feedback, 2005). A probable reason for the minimal dissent was because the assessment tools are diagnostic in nature with the explicit purpose of helping teachers better understand the students' numerical thinking (Ministry of Education, 2004b). With that purpose being the primary driver of the initiative, teachers felt comfortable with aggregating the diagnostic information and reporting effect size gains for students' positive movements through the number framework stages. This situation reinforces the arguments made in Chapter 4 about the prominence of formative assessment as a lever for improvement in the minds of New Zealand educators.

7.1.3. The third-space community of practice

The third space community of practice activated the decisions made by the collaboration in the fourth space. Its membership was more straightforward than the design and development team in that it involved learning connections between
professional development facilitators and teachers. The facilitators had to break away from their generic traditions and work with teachers in their context to change their thinking. Of particular importance was the need to break the teachers' preference for written algorithms to solve mathematical problems. Learning connections to do so centred on observing, modelling, talking and thinking about the content and pedagogy used in mathematics lessons (Ministry of Education, 2004b). The approach relates to the third policy principle mentioned in the previous section of constructing teacher knowledge in their own context of practice (Higgins & Parsons, 2005). Hence the connections occurred in and around classrooms. There was no one specific learning place, such as the staffroom or an off-site professional development seminar room, where practitioner learning is often assumed to be occurring. It was more a process of learning on the job.

Feedback from one policy developer attached to the project stated that the process was challenging, particularly at the outset of the professional development programme (National senior advisor for numeracy, New Zealand Policy Developers' Feedback, 2005). The facilitators had to confront teachers' prior beliefs and practices that were not consistent with the principles underpinning the number framework and diagnostic interview. Those principles centred on embedding into the minds of teachers four standard teaching practices that they should use back in the schools they worked in once they left the third space. Four standard teaching practices were synthesised from the various source documents:

(i) finding out what the students know and what strategies they use to solve number problems by using a diagnostic survey;
(ii) designing lessons that will address students' knowledge gaps and increase their repertoire of problem-solving strategies;
(iii) teaching the lessons in such a way that the teachers' and students' thinking is made explicit; and
(iv) checking that the lessons did what they were designed to do by using follow-up diagnostic surveys.

Researchers found that those practices successfully changed teachers' thinking about the importance of growing understanding before teaching algorithms (Higgins,
Chapter Seven: Policy-Research-Practice Collaborations in New Zealand

The change in thinking was well captured in teacher interviews conducted as part of evaluations for the initial trial of the numeracy project and once it was well underway. One quote, in particular, captures the shift in teachers' thinking: "I think, for little kids, it's probably really important not to introduce the algorithms until you're sure they've really got it because it's an easy way of working the answer out without actually understanding how the heck you got that answer" (Higgins, 2001). So the mind shift was more about understanding the process of solving problems than about inducting students into easy ways of producing the right response.

7.1.4. Summary of the vertical policy-research-practice collaboration

The description of the policy-research-practice collaboration operating within the Numeracy Development Project identified two tiers of learning connections that helped create a useful set of practices for teaching students to solve number problems and an equally useful set of tools to guide those practices. The standardised frameworks, lesson formats and resources developed by the fourth space community of practice determined to a large extent the nature of the learning connections between the facilitators and teachers in the third space connections. Contributions from the fourth space, however, were not produced by one dominant force such as the policy community or the research and development community as was the case in the vertically-driven international initiatives. It was a matter of the various groups making an input into the developments where they had relevant expertise. So they successfully broke out of their silos to work on the whole project rather than parts of it. In terms of implementation, the clear direction about relevant content knowledge and pedagogy from the fourth space helped the facilitators and teachers uncover and develop the students' thinking about solving number problems in a more consistent manner than had been the case in the past. Standing back to reflect on the two tiers of learning connections, it became apparent that the participants involved in the collaboration were accepting standardised ways of doing things in a non-standardised legal framework.
7.2. A Horizontal Policy-Research-Practice Collaboration

Learning connections between the officials, researcher-developers and practitioners in the Mangere learning network represent a horizontally-driven policy-research-practice collaboration. Figure 13 shows the connections and the outcomes from them. They reveal some interesting differences and similarities from a much smaller scale policy-research-practice collaboration than the one attached to the Numeracy Development Project. Once again, the diagram is best read from bottom to top through the trajectories of the three participating groups. They are listed at the bottom of Figure 16 from left to right as the researcher-developers, the policy developers (includes local officials and national policy developers) and the practitioners.

7.2.1. Pre-intervention development needs

Researchers tracking early developments in the SEMO project in 1998 and 1999 helped all three groups come to realise that what they were trying to do to improve schooling was not necessarily impacting positively on the underachievement problems and, in some cases, was adding to the problems (Robinson, 2000; Timperley et al., 1999). One quote, in particular, by Robinson (2000) captures the collective state of not knowing what to do. “We don’t know nearly enough about how to teach students from disadvantaged backgrounds to reach the goals that their families aspire to. When I say ‘we’, I mean university researchers, policy makers, parents, community leaders as well as educators” (p.40). So each group had to make some shifts from the past if they were going to make a useful contribution. In brief, the pre-intervention development needs of the three groups included: researchers learning to develop formative relationships with practitioners rather than producing research disconnected from practice (Robinson, 1998); policy developers becoming more actively involved in the learning processes and thereby more knowledgeable about improvement processes after ten years of staying well away from schools (Annan, 1999); and the practitioners replacing data-free approaches with rigorous inquiries to check that what they were doing was making a positive difference for all the students (Timperley et al., 1999). Those developmental needs set crucial learning agendas for
the change process in the third and fourth space communities of practice operating within the Mangere learning network.

**Third space**
Researcher-developers and middle managers
- Engaged in mastery learning in domain knowledge: how to analyse and use student achievement data
- Co-constructed classroom interventions to improve students' reading comprehension

**Fourth space**
Researcher-developers, middle managers, principals and local officials
- Agreed on domain knowledge: analysing and using achievement information
- Developed a self-review tool and a cross-school achievement information management system

**Past**
Disseminated research findings through hand over encounters
Policy of leaving schools to identify and solve achievement problems

**Present**
Middle managers and principals helped teachers
- Agree on common assessment tools
- Analyse achievement problems together
- Adjust teaching practices to solve achievement problems
- Check that the teaching worked practices were successful

**Future**
Learning encounters with practitioners and policy developers
Evidence-informed policy
Evidence-informed teaching practices

Researchers
Policy developers
Practitioners (Middle-managers, principals and teachers)

Figure 13. The Policy-Research-Practice Collaboration in the SEMO Project.
7.2.2. The fourth-space community of practice

The purpose of the fourth space community of practice attached to the Mangere learning network served a similar purpose to the one in the Numeracy Development Project. It was a place for making important reform decisions and for developing reform tools. In this case, however, the membership was less complex that the membership of the fourth space in the Numeracy Development Project. It included a small lead group and a larger action group that got things going in the individual schools and classrooms. The small lead group had four members including a middle-manager chairperson, a researcher-developer, a lead principal and a Ministry official. The larger action group was made up initially of the middle managers from the sixteen schools participating in the network.

The thinking behind the middle managers assuming a leadership role in the fourth space on behalf of the schools was that they were best placed to create links between that space, the third space and what was happening in the schools (Ministry of Education, 2003d). They were directly responsible for formative supervision of the teachers and they were influential in senior management and board meetings because they attended those regularly. Principals were not as well positioned as the middle managers to take the lead role in the third space mainly because they were tied up in many other self-managing challenges that come up in what are considered large schools in New Zealand (350-500 students). That did not mean that the principals did not enter and actively participate in the fourth space. Attendance records in the minutes of the management team’s regular monthly meetings indicate that as the network was forming they tended to participate intermittently in initiative meetings, mainly as observers. However, as time went on the principals became interested in participating, as opposed to observing, once they realised that they needed knowledge about analysing and using student achievement data as much as the middle managers did to fulfil their instructional leadership roles in the schools.

A priority decision made by the middle managers and principals and their learning partners in the fourth-space decision-making arena was to learn and disseminate

---

34 The number of schools involved in the Mangere network has reduced over time. In 2000 when the network emerged, 16 schools in Mangere participated in a review of their planning and assessment processes (Ministry of Education, 2000). Eleven schools went on to participate in classroom interventions (Ministry of Education, 2002b) and now eight of those schools are developing the achievement management system in Mangere (Ministry of Education, 2003d).
knowledge about how to effectively analyse and use student achievement data to inform governance, management and teaching decisions. Splitting the priority knowledge development needs into two parts, i.e. ‘analysis’ and ‘use’, was intentional (Ministry of Education, 2000). It was a way of making sure that when school governors, managers and practitioners were planning to improve something, they identified the presenting achievement problems and carefully analysed them at the outset (‘analysis’). Then they were expected to ‘use’ the information to critique the current situation and to develop appropriate solutions to solve the presenting problems. Members of the fourth space were, therefore, not pre-determining lessons and resources to address a generic achievement problem as was the case in the fourth space in the Numeracy Development Project. In this case, it was a matter of refining planning and assessment skills so practitioners were better able to identify presenting achievement problems and to tailor their lessons and resources to address those problems. A recent publication which synthesises much of the work of the Mangere learning network in this regard suggests they were trying to develop a culture of inquiry (Robinson & Lai, 2006). By a “culture of inquiry”, they refer to a definition developed by Toole and Seashore Louis (2002); “a school-wide culture that makes collaboration expected, inclusive, genuine, ongoing, and focused on critically examining practice to improve student outcomes” (p.247). In their view, such collaborative learning communities, often referred to as “professional learning communities”, are not comfortable collaborations through which teachers merely share ideas but they are opportunities for rigorous investigations of schoolwide teaching and learning (Robinson & Lai, 2006, p.198).

One of the main reasons why the leaders of the Mangere network prioritised inquiry learning as a fundamental reform practice was because they were influenced by some key research findings supporting that stance. Explaining the way that happened shows how the researchers were creating learning connections with practitioners as distinct from handing them information. Initially, the researchers made some important discoveries. They found out that past efforts by many Mangere schools to plan for improvement via SEMO left out a careful analysis of the existing achievement problems (Timperley et al., 1999). As the schools went on to implement their plans, the researchers found that those interventions that did analyse and use achievement data typically made a positive difference (Robinson et al., 2000;
Conversely, those that were not analysing and using achievement information could not prove that what they were doing was worthwhile. Just because the researchers discovered these important findings did not mean that practitioners involved in the Mangere learning network were going to change their ways.

What the researchers did with their findings made a crucial difference. They initially fed back and discussed the findings through meetings with the network leaders (Mangere Learning Network Secretariat, 2001a) with the individual school leaders and also with Ministry officials (Ministry of Education, 1998-2002). Those rounds of critical discussions got the findings out among the members of the network but things did not stop there. The researchers then re-connected with leaders of the network to lift the findings to a higher theoretical level in order to publish them in international journals (Robinson, 2000; Timperley & Robinson, 2000, 2002). Those connections ensured that the leaders of the network had a deeper understanding of the findings and became advocates for doing something about them. The researchers then broadened their dissemination strategy again by helping a variety of practitioner and policy developer-participants to conduct evaluative research projects of their own that explored different aspects of the original research findings (Robinson, 2003; Robinson et al., 2001). Publications from those projects extended the information base about what constituted inquiry-based learning in school improvement in districts like Mangere. Those publications also added a deeper level of sense-making within the Mangere learning network about the original research findings. One principal who attended a launch celebrating the publication of the first round of projects stated, “we are very appreciative of the findings from these projects, which are reminding us to check that what we are doing is making a difference. It is so easy to just get on with what we think is the right thing to do and to not check” (H. Hampton, personal communication, October 12, 2001).

Even though the members of the fourth space in the Mangere network grasped the importance of inquiry-based learning faster than their colleagues in other cottage industry initiatives, they still struggled at the outset to get their agenda sharply focused on the domain knowledge (Mose & Annan, 2002). Karen Mose, the chairperson acknowledged that they easily slipped into discussing procedural and administrative issues. The problem was partially addressed in the fourth space by
Karen Mose, the chairperson, and the lead team transferring all administration matters to a separate administration meeting monthly and setting a learning agenda for their professional development meetings (Mangere Learning Network Secretariat, 2001b). However, it was not fully addressed until the researcher-developer, referred to by network participants as the problem analyst, joined the lead team in the network and added considerable depth to the learning programme. Karen said,

The problem analyst role has enabled a move away from the safety of administration within the meetings toward a more direct focus on teaching, learning and achievement. New learning opportunities were built into an area-wide professional development plan that focused on skills needed by managers to lead inquiry processes. Meetings have been restructured to allocate specific time component within the agenda identified as the "learning component". (Mose & Annan, 2002, p.8)

An important part of the problem analyst’s role in leading the area-wide professional development programme was setting the middle managers and principals mastery tasks and checking for understanding through mastery testing (Mangere Learning Network Secretariat, 2002b). A feature of the mastery learning programme in the fourth space was its mix of developing a technical understanding of the achievement information and acquiring skills and knowledge to work out the causes of the achievement trends. The technical aspects of the programme, such as content analysis of data and theory-informed use of data, helped the middle managers and principals make sure that the information they were collating within their schools was valid and reliable. It also helped them to more accurately present achievement information and to talk about it at staff, board of trustee and network meetings.

In terms of inquiring into the causes of the achievement trends, the problem analyst taught the principals and middle managers the technique known as problem based methodology (Robinson, 1993; Robinson & Lai, 2006), which I used in Part II to analyse the constraints surrounding the quality of evidence attached to the international and national initiatives. The network leaders found the technique to be particularly useful for analysing causal reasons for achievement trends (Mangere Learning Network Secretariat, 2002a). The problem analyst had considerable
expertise in helping the fourth space participants use this approach. She had been a lecturer of a postgraduate university course in research methods and the technique was an important part of the course. That expertise coupled with the demands of the mastery learning programme ensured that the network participants were not going to use the No.8 Wire approach to analyse achievement problems.

Useful reform tools evolved as members of the fourth space community of practice learned more about analysing and using student achievement data. When the members had limited knowledge at the beginning of the process in 2000, they developed an audit tool to review and improve their schools’ planning and assessment processes (Ministry of Education, 2000). Critical to improving those processes was the process of convincing at least half the principals and two thirds of the teachers that an evidence-informed approach was the way to go (Timperley, 2002). Although the researcher tracking the audit exercise indicated it made some in-roads into that problem, it at best helped develop a general understanding of effective planning and assessment processes (Timperley, 2004a). It did not identify specific parts of the processes that were causally linked to achievement problems. Additionally, there was no evidence to suggest the exercise raised students’ academic achievement levels. So with limited knowledge to do anything more sophisticated, members of the learning network appeared to fall back on the No.8 Wire approach to make a start in this new instructionally-focused phase of work.

As they became more knowledgeable and had the expertise of the problem analyst at hand, the next tool they developed was a far more sophisticated undertaking than the development of the audit tool. That tool was the achievement information management system described in the scenario in Table 8 (4.2.2.2. p.99). It was intended to be a user friendly tool to help intervention teams in the third space and practitioners in their school-based syndicate teams to identify and to solve context-specific achievement problems. Without repeating the scenario in Table 8, the key point here is that the more knowledge members of the fourth space community of practice had, the more they demanded of themselves considerable rigour in the development process of their second tool. A positive consequence of the rigour was the greater ease with which achievement information could be used to design, monitor and publish the results of the Year 4-8 students’ reading comprehension intervention (Lai et al., 2004; Lai et al., 2005; McNaughton et al., 2004).
Chapter Seven: Policy-Research-Practice Collaborations in New Zealand

The rigour in the fourth space also meant that the vignettes that local officials were able to pass on to national policy developers became real achievement trends. They were not passing on process-oriented stories that so often act as proxies for achievement information. They were passing on a steady flow of achievement information (Ministry of Education, 2005b). A case in point was a report of Year 3 students’ reading comprehension results from three schools participating in the learning network at the beginning and end of the 2003 school year based on the Supplementary Tests of Achievement in Reading (STAR, Elley, 2001) test results (Lai, 2004). What was so interesting about this development was that the fourth space leaders were more than comfortable for the Ministry’s senior officials and Ministers in parliament to receive reports about the achievement trends in their schools. It suggested that the members of the Mangere learning network trusted the government officials to only use the information to further develop the learning relationships between the policy, research and practice communities. The fact that government officials have not attempted to publicly embarrass the schools or the disadvantaged communities from the steady flow of achievement information from the network justifies the trust that has developed.

7.2.3. The third space community of practice.

Third space communities of practice in the Mangere learning network provided the middle managers and principals an opportunity to design classroom interventions with their preferred researchers-developers and to help the teachers they worked with to implement them. The design process involved two steps (Ministry of Education, 2002b). The first step was for the principals and middle managers to move out of the fourth-space community-of-practice to find a group of local schools to work with who had a common interest in solving a particular part of the overall achievement problem. Two examples of common interest were improving reading comprehension in mainstream classrooms and improving the delivery of programmes in Pacific bilingual classrooms (Ministry of Education, 2000). The second step was for the lead practitioners from the groups of schools to contract a preferred researcher-developer to help them design their interventions. For instance, eight schools in Mangere interested in improving reading comprehension contracted researchers from the Woolf
Fisher Research Centre with expertise in literacy interventions to help them. Meanwhile, five schools with Pacific bilingual classes contracted an education advisor who had strong links with the Pacific communities to clarify the purpose and the effectiveness of the bilingual programmes.

Because the principals and middle managers were the contracting agents and keen to be involved in the design process, the contractors gave their practitioner knowledge and understandings equal status to their own. In a relationship with limited power and control issues, the researcher-developers and the middle managers in the third space were able to co-construct classroom interventions and implement them.

Representatives of the Mangere learning network captured the development principle of authority belonging to the group as a collective unit of inquiry.

In the study reported here, the collection, analysis and discussion process took place in the context of collective analytical and problem solving skills where teachers collaborated with researchers and professional developers to co-construct the professional development aimed at sustainable improvements in student achievement. This assumes teachers have professional expertise distributed within and across schools and are able to contribute as co-participants in a research-based collaboration. (Lai et al., 2005, p.6)

To help teachers implement the interventions designed in the third-space communities of practice, the principals and middle managers and their preferred researcher-developers had to fulfil several roles. They had to move from a co-design role to become trainers and supervisors of the teachers.

The two most effective interventions were those that set out to improve Year 1-3 and Year 4-8 students' ability to read and to comprehend what they read (Lai et al., 2004; Lai et al., 2006; Lai et al., 2005; McNaughton et al., 2004; Phillips et al., 2001; Timperley, 2002, 2004a; Timperley, Phillips et al., 2003). Both of those interventions built on the approach taken in the original successful intervention designed specifically for teachers of Year 1 students (Phillips et al., 2001). Central to that intervention were four standard practices that represented an inquiry-based learning approach that the practitioners in the network agreed to use in the subsequent interventions. The standard practices which are listed in the oval on the right of Figure 13 were: (i) to use common assessment tools; (ii) to analyse the achievement
information to identify priority achievement problems; (iii) to locate teaching practices related to the achievement problems and adjust them; and (iv) to check the impact of the adjusted teaching practices on the problem through follow-up assessments. Those four practices are similar to the common practices developed through the Numeracy Development Project. Both inquiry processes started by checking what the problem was then encouraged the use of specific teaching practices to solve the problem and then checked for success before moving on. Gains in achievement were not assumed.

A feature of the third-space activity in the effective learning networks was practitioners’ critique and challenge of each others’ professional decisions and actions. One study identified a particular type of talk that emerged in the SEMO-related networks that is being particularly helpful in this regard (Annan et al., 2003). The particular type of talk of interest is called “learning talk”. It was discovered by taping, transcribing and analysing the talk that a group of teachers were using in their professional conversations within one of the Year 4-8 research-practice collaborations. The analysis task separated out all the talk that was not relevant or peripheral to solving achievement problems and revealed the talk that was most relevant. Three categories of talk emerged from the study as the most useful; analytical talk, critical talk and challenging talk. The study defined those three categories as “talk that analyses the impact of teaching practices on student learning is analytical talk; talk that evaluates the outcomes of that analysis is critical talk; and talk about making changes to ineffective ones is challenging talk” (Annan et al., 2003, p.31).

We developed an argument in that study that the three categories of learning talk are inter-related. In summary, the argument was that challenging talk about ineffective practices is suspect unless there has been critical talk that judges the practices to be ineffective. Those judgements are, in turn, suspect without analytical talk. That talk concentrates on finding evidence that supports the judgements and rules out other judgements. Combined, the three categories of learning talk became central to the inquiry-based approach that the network used to design, implement and review the success of their interventions.

Reflecting on learning talk as a tool for improving inquiry skills and knowledge across a group of schools, I believed it is a useful way of supporting practitioners to make sense of many important international and national research findings. It avoids
Chapter Seven: Policy-Research-Practice Collaborations in New Zealand

hand-over encounters. For instance, the ‘analytical’ component created a demand for achievement information to become the primary agenda item of professional learning meetings focused on raising achievement (Timperley, Phillips et al., 2003). That component of the language also helped practitioners to develop knowledge and skills in problem analysis (Robinson, 1993; Robinson & Lai, 2006) and to avoid the problem which was identified early in the SEMO project of adding on programmes (Timperley et al., 1999).

The ‘critical’ component encouraged practitioners to search for evidence that supported their beliefs about the causes for low achievement (Spillane et al., 2002; Timperley & Robinson, 2002). It also encouraged them to search out discrepant data that might cause them to change their beliefs. A case in point was the widespread discussion during the first reading comprehension intervention about the practitioners’ deficit thinking with regard to the capabilities of disadvantaged students and their families causing the students’ reading problems (Timperley & Robinson, 2002). Critical discussions about the success of the programme caused many of them to realise that it was their teaching capabilities that were playing a major part in the academic success or failure of their students.

Finally, the ‘challenging’ component helped practitioners to act differently and acting differently is recognised as the most difficult component of inquiry learning (Spillane et al., 2002). Considerable traction appeared to be made in this regard within the Mangere learning network once the problem analyst helped the participants review their annual professional development programme. The idea was to ensure that each annual plan triggered more effective professional actions in the subsequent year. For instance, the 2003 review (Lai, 2003) stated that the inquiry skills that practitioners had learnt to use to improve the students’ reading comprehension could be successfully applied in other content areas, such as mathematics and science. The problem analyst argued for generalising the inquiry learning approach based on two pieces of evidence. One piece of evidence was the successful in-school case studies of lead-practitioners in the Mangere learning network showing how the AUSAD skills could be successfully generalised to curriculum content areas other than literacy. The other was international literature which indicates that developing consistent inquiry skills to achieve area-wide coherence is a crucial variable in raising student achievement (Newman, Newman, Smith, Allensworth, & Bryk, 2001).
To bring this to a close, learning talk in the Mangere network helped bring alive research findings about the centrality of achievement information for schooling improvement meeting agendas, the avoidance of add-on programmes, the need to have evidence-informed beliefs and the need to check that reform practices are actually in use. My main point is that the publication of these critical findings in books and articles is an important milestone in professional learning but it is not by any means the end point of an improvement journey. The findings have to be talked into existence, reported by those who they were intended to help and debated as implementation difficulties arise and new findings become available.

7.2.4. Summary of the horizontal policy-research-practice collaboration

Once again, the findings of the policy-research-practice collaboration operating within the Mangere learning network provide evidence that a more standardised way of doing things is taking hold within the non-standardised legal framework. The analysis identified two tiers of learning connections that helped develop a useful set of standardised practices, as did the collaboration attached to the Numeracy Development Project. In this case, however, the lead group in the fourth space community of practice were not directive about the sorts of teaching practices and classroom resources that should be used. They were more interested in developing stronger planning and assessment processes. So the primary change principle in the fourth space collaboration in this case was to influence the way that third space communities of practice went about designing and evaluating their interventions. That change principle contrasted with the more directive approach used by the fourth space leaders in the Numeracy Development Project as to what and how practitioners were to teach students to solve number problems.

Greater flexibility in the third space for researcher-developers and practitioners to design and develop interventions did not mean that the practitioners ended up with any more autonomy than those involved in the Numeracy Development Project. Once their leaders had co-constructed the interventions with their preferred researcher-developers, teachers were closely supervised by their middle managers. They also kept a much closer eye on each others’ practices and results than they had in the past. Meanwhile, the local officials watched and learned from the third-space interventions.
with interest to pass on relevant findings to their national colleagues for consideration in their important policy work to improve the system overall.

7.3. Summary and Conclusions

The investigation in this chapter provided more in-depth information than that in the previous chapter to answer the research question that asks about development and implementation characteristics relevant to accelerating the learning of effective reform practices. This more detailed discussion centred on the learning connections within the policy-research-practice collaborations attached to the Numeracy Development Project and the SEMO project. Of most interest were those connections that transformed useful theoretical and practical ideas into effective instructional practices and resources. The national collaboration attached to the Numeracy Development Project relied on a set of vertical learning connections and the localised collaboration relied on horizontal connections learning dimension. Yet they both developed a similar set of standardised inquiry-based learning practices for practitioners which focused directly on instructional improvement. So the collaborations, regardless of where they were situated, set up a continuous cycle of assessing students' academic progress then teaching to address identified needs and then re-assessing them to make sure things had improved.

Both collaborations evolved a mix of connections without any particular group involved, i.e. national policy developers and/or local officials, researcher-developers and practitioners, dominating proceedings. The evolutionary process in both the vertical and horizontal learning dimensions indicated a preference for distributed expertise in schooling improvement in New Zealand. That preference avoided the development of centres of expertise but brought with it a complex mix of learning connections within the systemic hierarchy and within localised learning networks. In contrast, there was an international preference for expertise to reside in the policy community (in England) or the research community (in the United States) in centres of expertise. The international preference led to more straightforward learning processes and connections at different points in the systemic hierarchy between the centres and the schools. That is not to say that the more straightforward international
processes and connections were any less sophisticated than the more complex ones in New Zealand. Additionally, they were no less successful.

Another interesting finding about the desire to distribute expertise in the two collaborations in New Zealand was the Ministry's desire to take a back seat position. National policy developers and local officials were trying to position themselves and their policies underneath the research and development and practice communities. They indicated this preference implicitly by staying in the fourth-space communities of practice one step removed from the front-line intervention work in the third-space communities of practice. That appeared to be a deliberate strategy to push for leadership of the development and implementation of school improvement within the education sector and not within the Ministry. This strategy is consistent with the same long game used to try and spread expertise in evaluating schooling improvement initiatives across the system.

To conclude this chapter and Part III as a whole, the sophistication of the learning systems and the various combinations of learning connections and collaborations highlight at least two interrelated theoretical assumptions which underpin useful school reform learning processes. One assumption is that the task of learning effective school improvement is extremely difficult. The second is that no one organisation or role can transform the available knowledge into relevant reform practices and resources. These assumptions indicate that New Zealand's No. 8 Wire cultural norm of just getting on with job of solving problems is not a satisfactory development or implementation characteristic to accelerate the learning of effective reform practices. The fourth and final part to this thesis builds on the two theoretical assumptions underpinning the discoveries made in the investigations in Part III to develop a theory of connected and consistent learning for instructional improvement.
PART IV: Consistency and Connectivity to Improve Instructional Practice

PART IV

Consistency and Connectivity to Improve Instructional Practice

Purpose

Part IV of this thesis is intended to offer some theoretical and practical ideas to policy developers, scholars and practitioners who are considering how to accelerate the learning of effective school reform practices. Three research questions focused this thesis onto solving the problem of the slow pace with which this learning is currently occurring in New Zealand. Solving that problem is expected to contribute towards solving the overarching problem of underachievement among disadvantaged students. The first research question asked what were the most effective national and international examples of school improvement. The second question asked about the condition of the evidence base for making claims of effectiveness. The third research question asked what can be learned about developing and implementing effective school improvement from those national and international examples. Ideas presented in this section, which were generated from the investigations into those three questions in Parts II and III, are intended to inform those professional educators who are involved in various types of schooling improvement initiatives.

Approach

In Part IV, it is not my intention to advocate for one or more of the seven initiatives or for one of the three learning process models developed. What is
emphasised in this section, are the conditions that accompany accelerated schooling improvement. To review those conditions, I draw on the work of Pawson (2002) who argued that 'generative mechanisms' that lead to change are dependent on the specific contexts in which they operate. Researchers must generate 'transferable theories' rather than 'best buy' theories. Pawson (2002) states that it is a matter of working out that “This programme theory works in these respects, for these subjects, in these kinds of situations” (p.324).

This chapter begins with a two-part discussion. The first part of the discussion concentrates on the condition of the evidence base as that was used to identify the characteristics of effective school improvement approaches. The second part of the discussion centres on three characteristics that form a theory for connected and consistent learning, focused on instructional improvement. Implications of the thesis are then presented in terms of theoretical consideration for the general field of school improvement and practical ideas for New Zealand’s school improvement work. The chapter also considers the limitations of the thesis and the need for further research before concluding statements can be made.

Connectivity is a central concept in the theory of learning developed in this thesis. It is not a new phenomenon to help solve complex problems within and beyond education. The concept is an important lever for profit and service delivery advantage in the fields of business and telecommunications (Barabasi, 2002; Siemens, 2004; Wenger et al., 2004). Although some scholars have started exploring the potential of connectivity focused on instructional improvement in the field of school improvement (Coburn, 2005; Fullan, 2005a, 2005b; D. Hargreaves, 2003b; Higgins & Parsons, 2005; Lai et al., 2004; McNaughton et al., 2004; Spillane et al., 2002; Stein & Coburn, 2005; Veugelers, 2005), there is a great deal more to be done to make the theory explicit and to apply it well in practice. The nature and extent of connectivity surrounding instructional improvement discovered in this thesis and the discussion points made in this chapter about those discoveries contribute to that important work.

What the discussion does not attempt to do is to inform connections that do not involve professionals. The scope of this thesis (see Chapter 1, p.15) precluded an investigation of a broader range of connections between professionals and other groups involved in schooling, such as parents. However, it is important to remind the reader that research findings into those other types of connections are also extremely important to consider in thinking about accelerating their learning. For instance,
Bishop et al's (2003) findings from Te Kotahitanga in New Zealand about the importance of learning as well as caring connections between teaching professionals, Maori students and their families are well worth considering in light of the overall poor track record nationally and internationally in helping indigenous peoples succeed in school.
PART IV: Consistency and Connectivity to Improve Instructional Practice
CHAPTER 8
DISCUSSION AND CONCLUSIONS

The first part of the discussion centres on the findings from Chapter Three and Four that evidence of effectiveness was stronger in England and the United States than in New Zealand. A brief recap of that claim precedes a discussion of the two reasons for the difference. One reason is patterns of investment into research and development. The other reason is that outcomes-focused research is easier to produce in England and the United States than it is in New Zealand. Explanations for both reasons are presented and discussed to get to the heart of why the evidence is stronger overseas.

8.1. Condition of the evidence

Evidence attached to the five international initiatives received, the highest possible rating because there was sufficient robust summative information to make judgements about the magnitude of student progress. The information came from evaluations that were characterised by methodological rigour and a strong presence of third-party evaluators. Although there were different ways of achieving those characteristics in England and the United States, the result was the same. There was little doubt that positive academic gains were being made by the targeted students. In contrast, there was insufficient summative information attached to the initiatives in New Zealand to be sure about what was happening with all the students’ learning. There were a few evaluations with methodological rigour but insufficient to lift any one initiative’s evidence to a ‘strong’ rating. A major shortcoming of the evidence was the absence of third-party summative evaluations.

8.1.1. Patterns of investment into research and development

Investments into outcomes-focused research in England and United States made sure that funding was closely tied to high quality summative evaluations. There was an expectation in the funding allocations that the magnitude of improvement in student achievement would be measured and reported. Furthermore, third-party evaluators were expected to play an important part in validating claims of
effectiveness. Summative evaluations, therefore, were not negotiable. Contracting arrangements for evaluations of the English initiatives and the Comprehensive School Reforms, for instance, made the expectation very clear. An important part of the international research team brief for evaluating the English initiatives was to give independent feedback on progress against the challenging national standards (Earl et al., 2003). In the case of the Comprehensive School Reforms, it is hard to imagine how anyone associated with them would not be thinking about scientific outcomes-focused research given that two of the 11 federal funding criteria specifically require it (United States Department of Education, 2002) and multiple references are made to it in the No Child Left Behind Legislation (United States Government, 2002).

The requirement for quantitative measures of improvement did not exclude formative evaluations. To the contrary, the leaders of all five international initiatives enjoyed the benefits of formative feedback from researchers. Some received it from researchers who were also conducting summative evaluations and others received it from researchers focused purely on informing the initiatives. Researchers investigating the benefits of Comprehensive School Reforms note that information from summative evaluations can be just as useful for informing practice as it is for judging the magnitude of improvement (Rowan et al., 2004).

When fed naturally-occurring work groups such as grade level teams, school improvement teams, or other faculty groups, evaluative information on processes and outcomes directly addresses the concerns of innovation users about the practicality and consequences of their change efforts, and (if the process is working well) can enhance the commitment of school staff to the change process. (p.14)

The important point here is that the international investments into evaluative research were balanced between finding out how well the students were doing and helping the developers and implementers of initiatives improve what they were doing. This sort of balance is what at least one of New Zealand’s most credible educational critics has argued for (Nash, 2002).

In contrast, investments into research in New Zealand are not contingent on high quality evaluations so grades of quality are acceptable. One reason for the variable quality evidence is the number of small scale formative evaluations. Those evaluations constitute research and development partnerships. They are attached to
some but not all the cottage industry initiatives and they are now featuring in national initiatives like the Numeracy Development Project. Investments into formative evaluation have not precluded some investment into summative research. However, that investment is small based on the limited return from the search for evidence of effectiveness in Chapter 5. Additionally, a lack of independent third-party evaluation diminishes the value of the small amount of summative information that has been produced. So the evaluation of New Zealand’s school improvement initiatives is weighted much more heavily towards formative assessment than in England and in the United States.

There are at least three possible explanations for the preference for small scale formative evaluations in New Zealand. They include, (i) a policy preference for local capacity building in both evaluation and development, (ii) a complementary policy preference for summative information to come from national and international surveys and, (iii) there is simply not enough money to do any more evaluation. A few details about these three explanations help to ascertain the most plausible one.

The first possible explanation of local capacity building is tied to an assumption that formative evaluation is integral to development, hence research and development partnerships based on formative types of evaluation. Evidence for this claim relates back to the findings in Part II (4.2.2.1) about the influence of the national assessment policy on school improvement in New Zealand, i.e. formative relationships between teacher and student not only help define achievement but they are a vital mechanism for improving achievement (Ministry of Education, 2001). It is also consistent with principles of local development and self-review inherent in the broader self-managing legislation (New Zealand Government, 1989). A consequence of the local capacity building is that it promotes certain types of controls over evaluations. Researchers are constrained to investigate what others in the partnerships consider to be priorities. Consequently, rather than summative evaluations being seen as an imperative for success, they remain a negotiable item with powerful partners who might have little or no understanding of programme evaluation.

The second possible explanation for the prevalence of small-scale formative evaluations in New Zealand relates to the relative policy silence about evaluating the effectiveness of schooling improvement initiatives across schools. There is explicit guidance in the national assessment policy for tracking achievement trends across schools but it is not tied to schooling improvement interventions. It relates to national
policy developers' use of achievement information aggregated from national and international surveys (Ministry of Education, 2001). Those surveys produce ample evaluative information for analysis but the problem with it is that it does not relate to all targeted students involved in specific school improvement initiatives. Additionally, researchers of school improvement initiatives cannot rely on the evaluative focus or the data collection timeframes of the surveys matching their requirements. Those problems make the national assessment policy largely irrelevant to evaluating the success of school improvement initiatives.

Its irrelevance does not mean there is a complete policy silence on the matter of evaluating the success of interventions across schools. Recent commitments made by senior policy developers indicate that the silence has been broken (Eppel & McGregor in New Zealand Policy Developers' Feedback, 2005). Not only have they made commitments to support outcomes-focused evaluations of national professional development programmes and cottage industry initiatives, they also endorsed the overarching evaluation framework for the raft of interventions supported by the Ministry (Ministry of Education, 2006c). However, the commitment is so recent that it is hard to read any policy directions into what is happening. That said, reflections from the Ministry's national manager overseeing the funding and monitoring of cottage industry initiatives are encouraging: "we were slow to get started but we are learning and we will get there" (National manager of the monitoring and support division, New Zealand Policy Developers' Feedback, 2005).

The third possible explanation for small scale evaluations is that a more sophisticated evaluation programme is simply not affordable. New Zealand is a small country compared with England and the United States and there are considerable constraints on vote education money. As part of the self-managing policy, a great deal of funding is given to school boards to set and address their own improvement agendas. Appropriate further funding for additional initiatives is given very careful consideration by government and whatever is allocated is considered extremely sensitive by Ministry officials. By sensitive I mean it must be spent for the purpose it was given. Decisions to evaluate some, and not all, of the cottage industry initiatives, for instance, were partly due to cost sensitivities. The idea was to evaluate a few of the initiatives in districts that were causing most concern and use the findings to inform developments in other districts.
Of the three explanations for small scale evaluations in New Zealand, the least plausible is the latter affordability argument. That aspect can be rectified by creating a high expectation that continuation of school improvement initiatives is contingent on high quality outcomes-focused evaluation. Such an expectation will inevitably cause some thinking about adjusting ratios of expenditure into development and evaluation and, more specifically, into formative and summative evaluation of schooling improvement initiatives. That is a prioritisation argument and not an argument for more money.

The most plausible explanation is a combination of the other two arguments; a preference for formative assessment to assist rather than assess the value of developments and the relative silence of national policy on cross-school evaluations that are useful for tracking the effectiveness of school improvement initiatives. Further exploration of those preferences is woven into the explanation in the next section about the greater ease by which outcomes-focused research can be produced in England and the United States compared to in New Zealand. The explanation focuses on the law, the Ministry’s theory for intervention and an avoidance of mistakes made overseas.

8.1.2. Ease of producing outcomes-focused research

Outcomes-focused research is easier to produce in England and the United States than it is in New Zealand because achievement information is more available for evaluation at the school level in those two countries. The international investigations in Part II and III provided evidence that there are systemic achievement databases tied to schools in England and the United States producing achievement information that school improvement researchers can access. Some researchers tap into data bases developed for specific initiatives (England, SFA, and NDP) and others tap into data bases for city and state-wide testing programmes (DI & NY). New Zealand researchers, on the other hand, only have one systemic data base to tap into that identifies schools (Ministry of Education, 2006d). It manages examination information (NCEA) for students in the last three years of schooling. There are no other systemic data bases for tracking students’ achievements in the earlier years of schooling apart from data in the national and international surveys, which are not suitable for evaluating school improvement initiatives. In the absence of systemic
data bases in New Zealand, school improvement researchers face a laborious task of gaining permission to gather data, gathering it, and, in many cases, cleaning it up before using it.

Plausible explanations for the national-international variations in the availability of achievement information hinge on the amount of direction given from the centre. Far less central direction is given to managing achievement information in New Zealand. I argue that three factors among a number are the main contributors to this situation. The first is that it is out of necessity. The law constrains central direction and those at the centre are obligated by law to expect school boards to manage student achievement information (Ministry of Education, 2006f), which means the boards make decisions about its availability to school improvement researchers. The second factor is the Ministry's minimalist theory for intervention. Despite the law, the intervention framework is now sufficiently developed to provide greater central direction. However, the Ministry's approach has been to encourage leadership in the sector. A critical part of that approach has been to continue developing the national education and administration guidelines so that schools can be held accountable for their developments (National manager with oversight for cottage industry initiatives, New Zealand Policy Developers' Feedback, 2005). Groups of schools interested in developing cross-school achievement management systems are encouraged to lead the development themselves. Even though this has proven to be extraordinarily challenging in the district of Mangere (see Chapter 5), central direction has not increased. The third factor is to avoid the traps associated with large scale achievement management systems overseas, such as teacher dissent, teaching to the test and public humiliation of schools. Combined, those three factors are constraining any more direction coming from the centre.

There are at least two consequential problems associated with minimal central direction in New Zealand. One problem is that it is taking a long time to develop high quality evidence of effectiveness and that is causing a moral dilemma. Disadvantaged students are the recipients of school improvement initiatives that may or may not be helping them achieve better results. The second problem with minimal central direction is that it is difficult for the centre to learn about whether the overall approach is working or not. Feedback from senior policy developers indicate that both problems have been acknowledged in a discussion document about developing an overarching evaluation strategy, which is yet to be released publicly (National
The same two consequential problems cannot be said for England and the United States. An outcomes-focused orientation and centralised learning are already in place. The learning process models in Chapter 6 accentuate this point. In the two international models in Figures 6 and 7 there are centres of expertise and outcomes-focused feedback loops to the policy contexts to monitor the overall approaches. There are no such centres in the model for the New Zealand initiatives and feedback loops to the policy context are conditional on the extent to which initiatives have any outcomes-focused information to offer. Instead, a feature of the model for the New Zealand initiatives is distributed responsibility for learning about what is going on. What appears to have come with this distributed approach is leeway for confidence within individual initiatives without benchmarks to justify it. That may be useful for creating a good feeling about being involved in a school improvement initiative but it allows those involved in less effective initiatives to think that they are doing equally as well as others involved in more effective initiatives. That sort of problem indicates that there is a price to be paid for the limited centralised learning from achievement information. The development of an overarching evaluation strategy may help address that sort of problem as long as the intent is to achieve common understandings and expectations about programme evaluation and to develop robust criteria for judging the quality of evidence.

8.1.3. Summary of the condition of the evidence

The condition of evidence of effectiveness was found to be stronger in England and the United States than in New Zealand because of a more balanced investment into formative and summative research and the greater availability of achievement information for researchers to conduct outcomes-focused research. The process of developing the discussion revealed to me that a preferred evaluation practice in New Zealand is to assist school improvement initiatives rather than assess their effectiveness. The practice draws on a strong preference within New Zealand’s education community for formative assessment over other forms of evaluation. That
Chapter Eight: Discussion and Conclusions

preference is not to say that summative evaluations of schooling improvement initiatives are not underway. Policy is being developed to start conducting routine evaluations of all interventions.

Crafting this section of the discussion also revealed to me that centralised learning about overall progress is constrained in New Zealand because of the limited availability of achievement information. Factors considered as main contributors to this situation are the law combined with a minimalist intervention approach by the Ministry and a fear of repeating mistakes made by others overseas. A negative consequence of the limited investments into summative research and the constrained centralised learning is that there is no way of knowing which schooling improvement initiatives need additional help to get better results. That leaves the door open for false confidence, which is something that New Zealand's school improvement movement may have to deal with when the evidence base does improve, as it is.

8.2. Development and implementation characteristics

This second part of the discussion begins by explaining the reasons for the variety of forms of effective school improvement found through the investigation in this thesis. Following that explanation, a set of interdependent common characteristics is revealed that forms a foundation for the theory of consistent and connected learning focused on instructional improvement.

8.2.1. Forms of effective school improvement.

Despite the apparent narrow focus of the definition of effective school improvement developed at the beginning of this thesis, it helped identify a variety of forms of effective initiatives. There were mandated scripts with tightly disciplined implementation strategies, a negotiated district-wide plan with local officials closely monitoring developments and a collaborative teaming and networking arrangements within and across schools. Of a range of different explanations considered for the variety, two of the most plausible were (i) the role of government influenced different forms of initiatives in different countries, and (ii) the developers' different theories for improvement led to the variety.
There were matches between the governments' roles and the forms of the initiatives that account for the variety. Government roles in school improvement can be active or passive. An active role can manifest itself through direct intervention by government agencies or indirect intervention by encouraging non-government agencies to fulfil an active role. A passive role allows non-government agencies to sort things out for themselves. All of the governments chose to fulfil an active role over a passive one which indicates that they all felt some responsibility for helping solve the underachievement problem. The nature of their active roles varied and as a consequence so too did the form of the initiatives.

Governments seem to have a particular influence over the process of generating, transferring and applying the reform practices. To recap, two processes were used. One was an importation process that relied on either a machine bureaucracy or a professional bureaucracy to make it work. The professional bureaucracy gave more leeway than the machine bureaucracy for practitioners to decide what and how to do things in the classroom. The other process was localised development via communities of practice. There is a match between the roles of the three governments and the processes used by the initiatives to import or locally develop the practices.

In the case of the English initiatives, the role of government and the importation process were one and the same thing. The English government's alignment of its educational agencies and policies and direct leadership role created the machine bureaucracy that drove the most widespread importation process. In the United States, federal government's indirect accreditation role and a direct leadership role by local officials in New York District #2 contributed to an interesting mix of importation and localised development processes. In New Zealand, government's direct active role via national policy developers and local officials influenced the formation of policy-research-practice collaborations to import a set of practices in the Numeracy Development Project and to develop a similar set of practices locally in the SEMO project. A feature of the government role was that it provided proactive leadership to develop and sustain design level collaboration and then stepped back into the shadows as research-practice collaborations intervened into classrooms.

---

35 Machine bureaucracies importing scripts (Direct Instruction & Success For All) as well as a localised development through teaming (The School Development Project) emerged out of federal government's stringent Comprehensive School Reform criteria. Additionally, a localised professional bureaucracy evolved in New York District #2 to help schools develop and deliver on tailored improvement plans to meet national standards.
Those matches between government roles and form help account for the variety. However, that explanation on its own would suggest that non-government participants involved in effective school reforms fall into line with official directions. That may have been the case in England and, to an extent in New York, but it was not the norm across the rest of the initiatives. In the others, researcher-developers’ design and implementation theories for improvement also influenced the form of the initiatives. Indirect government influence may have helped position researcher-developers to fulfil a lead role, but it was their theories that were put into use and they varied. For instance, Comer’s design theory promoted holistic assessments of students’ developmental needs and his implementation theory relied on consensus decision-making across collaborative teams. In contrast, Slavin’s design principles were based on mastery and collaborative learning in literacy lessons and he preferred a blanket implementation approach. New Zealand’s initiatives added another dimension to design and implementation theory through lead practitioners gaining equal, if not higher, status than researcher-developers and officials in the different collaborations. So controls over design and implementation theories by a range of professionals contributed to the variety of forms of effective school improvement initiatives alongside government influence.

Despite the variety, three common development and implementation characteristics prevailed. They are (i) a sharp focus on improving instructional practice, (ii) standardisation of classroom lessons or processes for planning and assessment, and (iii) learning connections. Each characteristic is explained in turn before showing how collectively they can make a useful contribution to speeding up the learning of effective school improvement.

8.2.2. Focus on instruction

The first common characteristic was a sharp focus on improving instructional practice. All seven initiatives targeted the use of teaching as a direct lever to raise student achievement. The School Development Programme was criticised as an exception to that rule by Cook et al. (1999), but Comer stated that instructional improvement was a vital part of the ‘Comer’ school-wide planning process. His defence indicated that schools that were not targeting the classroom as a point of intervention were not following the Comer programme (Comer & Haynes, 1999).
One explanation for the sharp focus on instructional improvement is the influence on the developers of their own experiences that pointed to instructional improvement as an important focus. Some of those experiences and exposures were from their involvement in interventions in the past and others were those attached to the current intervention. The developers were also exposed to a vast amount of literature about teaching interventions that have led to general accord among educationalists that teaching is the best systemic lever for improving student achievement (Alton-Lee, 2003; Ball & Cohen, 1999; Darling-Hammond, 1997; Elmore & Burney, 1999; D. Hopkins, 1987).

Another complementary explanation for the sharp focus on instructional improvement is a global movement within the field of school improvement away from trying to get broad coverage of something that might work. The movement is towards achieving greater depth in understanding and in the use of interventions that have been proven to work. Instructional improvement has proven to work but not in isolation to other factors (Alton-Lee, 2002; Coburn, 2003; Coburn et al., 2005; Hill, 2001). The theory is that it is better to focus deeply on instruction and its links to other factors, such as the nature of teacher talk and the quality of professional development programmes, than to treat instruction in isolation.

Those two explanations combined provide a strong rationale for focusing on instructional improvement. They reinforce Elmore’s (2005) argument that there is an instructional core that can be improved and that the school and system have to develop the capacity to enhance and support the resources, knowledge and skills of the teacher and student. His argument narrows the focus of school improvement onto teacher and student learning yet distributes responsibility for improvement across the education system: “This means literally using the school and the institutional structure as a mechanism to deliver resources and supports to teachers and students to enhance their learning. A controversial idea: that schools should become places dedicated to adult and student learning” (Elmore, 2005, p.222).

A sharp focus on instructional improvement on its own, however, is not what made the initiatives effective. It is quite possible to have a sharp focus on improving what

---

36 Comprehensive School Reform researcher-developers built their instructional focus on their previous experiences and involvement in related research and the developers of the English initiatives relied on findings from previous successful trials of the practices to develop their sharp focus.

37 Formative research findings in New York District #2, the Numeracy Development Project and the SEMO project pointed to instructional improvement as a priority focus.
happens in the classroom and not make any difference at all. There is a difference between the rhetoric of improving instruction and actually making instructional changes that impact positively on student learning. This is where standardisation comes into play.

8.2.3. Standardisation.

The second common characteristic was the creation and use of a set of standardised practices. Some initiatives standardised teachers' instructional content, teaching practices and classroom resources. Others chose to standardise some of the planning and assessment processes that influence teachers' decisions about what and how to do things in their classrooms. Assumptions about professional performance underpinned these two approaches. Standardising what happened in the classroom lessons assumed that the instructional core needed considerable improvement. Whereas standardising some of the processes surrounding the classroom but leaving practitioners to alter what happened in the classroom assumed that the professional core was satisfactory.

What was so interesting about standardisation was that it manifested itself in highly constrained environments as well as in environments with considerable freedom. For instance, standardised ways were inevitable in England with the government's determination to align everything from national policy to classroom practice. In contrast, the practitioners' high level of control over developments in the learning networks attached to the SEMO initiatives meant that the reform practices could easily have become non-standardised. The self-managing legislation (New Zealand Government, 1989) encourages a non-standardised environment yet the practitioners chose to standardise certain practices themselves. Readers may recall from Chapters 7 and 8 that the lead practitioners used a trial and error approach to agree on four standardised practices to improve their cycles of planning, implementing and evaluating their interventions. They ended up committing to highly disciplined standardised ways of operating, albeit it took them considerably longer.

38 English Literacy and Numeracy strategies, Direct Instruction and Success For All
39 New York District #2, the School Development Programme and the SEMO initiative
40 The four practices were (i) agreeing on and using common assessment tools, (ii) analysing the achievement information to find priority problems to solve (iii) teaching to address the identified achievement problems and (iv) checking that the teaching solved the problem by re-assessing the students.

208
than other initiatives to get there. Common across the variety of environments was strong commitment among support agencies and schools to the standardised ways of operating. That unification of effort is now well recognised as a critical improvement principle (Coburn, 2005; Datnow, 2005; Fullan, 2005a; Rowan et al., 2004).

Three factors in particular seem to have contributed to the use of standardisation as an effective improvement lever. The first factor is the realisation by the developers and implementers that a standardised approach gets better results. Over time, the developers and implementers of the initiatives learned that a particular set of instructional practices delivered better results. So they committed to it. The second factor was that standardisation is a useful way of assuring the integrity of the programme. It was not good enough to get by on the general principles of the programmes or to slip away from certain parts of the programme. Standardisation kept implementation tight so that the practices became ingrained into everyday life. A third factor was to urgently deliver on a political imperative to address what the incoming 1997 government called the long-term neglect of disadvantaged students (Earl et al., 2000). That factor was very explicit in England and it was also implicitly sitting behind the tightening of national standards in the United States.

In terms of sustainability, although political imperatives are heavily influencing standardisation in some school improvement initiatives at the moment, they are not necessarily going to be there to move things along in the future. The democratic right of developed countries' governments to prioritise what gets on the social policy agenda means that other priorities could take the place of school improvement in the future. The prioritisation of school improvement as a high priority agenda item is almost two decades old now and there is no guarantee that other agenda items will not displace it. What seems to have a better chance of enduring is school reform groups settling on and getting buy-in to a set of evidence-informed standardised practices and working hard to assure the integrity of their implementation.

Although that task is known to be challenging (Fullan, 1999; Stringfield & Datnow, 1998), prominent researchers in the field are suggesting it holds the greatest hope of a more successful future for school improvement (Datnow, 2005; Rowan et al., 2004). One study, for instance, found that 5 from 13 schools managed to continue implementing their preferred comprehensive school reform practices after three years of state and district-level policy changes (Datnow, 2005). That study concluded with two caveats for enduring success. One caveat is that initiatives leaders have to help
schools to deal with changing political demands in order to safeguard the reform practices. The second caveat is that affordable initiatives which are less demanding on the system have a better chance of survival in the face of policy changes.

To close off this section, standardisation is something that many educators appear to be wary about but constructed the right way, it is proving to be a valuable lever for improvement within the most effective school improvement initiatives. That is not a push for standardisation to be promoted as a solution for educational problems beyond solving the underachievement problem for disadvantaged students. The whole system should not be driven by effective solutions designed for specific groups. Conversely, where something does work for a particular group it should be systematised in such a way that it is protected for that group.

8.2.4. Learning connections

The third common characteristic was the fundamental importance of learning connections to transform the standardised practices from theory into practice. To recap, learning connections are the collective pursuit of knowledge among participants of school improvement initiatives to transform useful theoretical and practical ideas into effective practice. Vertical and horizontal learning connections were found through the investigation. The vertical ones were integral to the importation process used in vertically-driven initiatives and the horizontal ones helped the transformation process through communities of practice.

There are at least three reasons why learning connections are so critical to the transformation of effective theoretical and practical reform ideas into practice. The first reason is because there are three parts to the transformation process and all three need to be connected in order for the process to succeed. The three parts are: (i) the generation of relevant knowledge and ideas; (ii) the transfer of the knowledge and ideas into a useful set of practices; and, (iii) the accurate application of those practices. If the three parts are not connected then there is obviously going to be a breakdown in the transformation process. If they are connected but there is no active learning at each connection point, then the transformation is reliant on hand-over encounters. As I stated in outlining the research problem in the introduction, that alternative method has proven to be a main contributor to the slow pace of learning effective school reform (D. Hargreaves, 2003a). If the parts of the transformation
process are connected and active learning is only prevalent at some connection points and not others then practice is only likely to mirror the theory in part. So the intent of the change process is assured in the presence of effective learning connections and chance becomes a much greater factor without them.

The second reason why learning connections are so critical is that they help create feedback loops to make sure that the theory is in touch with the realities of practice. It is a reverse transformation process. It typically involves practitioners (appliers) providing feedback to the co-ordinators (transfer agents) and, in turn, the co-ordinators feeding back the practitioners’ views and their own to the developers (generators) to reflect on the improvement processes and the results. So learning connections help create a bi-directional flow of knowledge — theory informs practice and practice informs theory.

The third reason that learning connections are so critical is because they help resolve sticking points in the transformation process. The flow of knowledge through educational infrastructures is not the same as it is in other infrastructures where knowledge flows unimpeded through electronic airwaves or pipes (D. Hargreaves, 2003b). There are sticking points along the way. An example of a sticking point that learning connections helped to resolve in England was the LEA officials and consultants in England assisting practitioners to make sense of scripted professional practices in literacy and numeracy and discard their more autonomous old ways. School-based co-ordinators in the comprehensive school reform had to do likewise to help practitioners fit the standardised practices into their contexts. Discarding ineffective practices and unhelpful knowledge has been recognised as the most difficult knowledge management task to achieve (Spillane et al., 2002). Learning connections can help achieve that task alongside other more straightforward knowledge tasks, such as recognising and retaining existing effective practices and making alterations to existing practices that would be effective if a few changes were made.

All three reasons present a compelling argument for the importance of developing learning connections in effective school improvement. Without learning connections, the development of evidence-informed practices and consistent implementation of them is a matter of chance. Practitioners can learn and improve their practice from hand-over encounters but they do so on their own terms. Learning encounters create challenges to the independent thoughts of practitioners. They encourage debates
about the design and implementation theories as well as the way the practitioners are putting them to use. Taking the time to connect and have the debates, to come to resolutions and to act on those resolutions with integrity is part of doing effective school improvement.

8.2.5. Summary of characteristics

This second part of the discussion focused on development and implementation characteristics. The opening point was that there were several different forms of effective school improvement. Direct and indirect involvement of government and its agencies as well as the developers' design and implementation theories were offered as the most plausible explanations for the variations. The discussion then moved on to describe and explain three characteristics common to the various forms of effective school improvement. The three characteristics are a sharp focus on instructional improvement, the use of evidence-informed standardised practices and learning connections. They are not offered as stand alone levers for accelerating the learning process. Rather, they are considered interdependent. One without the other two is not going to achieve the same or better results than those achieved by the seven initiatives analysed in this thesis. To conclude, examples of effective school improvement do have characteristics that can help accelerate the learning process, although the next section reveals that even more can be done by building on those characteristics.

8.3. Accelerating the Learning of Effective School Improvement

The discussion points indicate that effective school improvement initiatives create a sharp focus on instructional improvement, help practitioners commit to and implement with integrity a set of evidence-informed standardised practices and, (iii) develop learning connections to make sure that what is implemented is effective and remains so. Replicating these three interdependent characteristics across initiatives that have little or no evidence of effectiveness is likely to speed up the process of learning effective reform practices. It will help bring the lower performing initiatives in line with what the high performing initiatives are doing. However, that act alone is not sufficient to solve the overall underachievement problem among disadvantaged students once and for all. Readers are reminded that the seven initiatives using the
three characteristics were only having a relatively low positive impact on student achievement. At best, therefore, a replication programme will mean a wider spread of low positive gains in student learning.

Consequently, other strategies are necessary to speed up the learning process. One useful strategy is to create learning connections across school improvement initiatives as well as within them. My theory is that learning connections are not only a useful lever for transforming sound theory into effective practice within individual initiatives but that they are also useful for linking initiatives operating in the vertical and horizontal learning dimensions. I am not talking about connections whereby the leaders of different initiatives meet together and share what they are doing. That is an add-on strategy, which is unhelpful in the field of school improvement (Timperley et al., 1999). The connections that I am referring to here are those that help a group of reformers advance a particular aspect of their practice. They are connections that help address problems that have been identified by a group of reformers or that have been inadvertently overlooked by them.

Evidence in support of promoting learning connections across school improvement initiatives as an accelerator of learning comes from the findings at the end of Chapter Six (6.3.3). That chapter inquired into the policy-research-practice collaborations in New Zealand and the way connections have started forming between the collaborations. The findings indicated that effective practices within cottage industry initiatives could help national professional development programmes solve problems and vice versa. To recap, a local-national connection involved the inquiry practices developed through the SEMO learning networks being adopted by the national literacy professional development programme after several years of getting no traction. A national-local connection involved the Numeracy Development Programme offering its set of evidence-informed practices to cottage industry reforms in the district of Manurewa as a better solution than developing an unproven local approach. Early results indicate that those connections contributed to better results (Ministry of Education, 2006i; Young-Loveridge, 2005).

The evidence is tenuous when put alongside the criteria for strong evidence of effectiveness developed in Part II of this thesis (p.51). However, it provides some positive feedback that developing a matrix of learning connections within and across the various vertical and horizontal levers is a useful way forward. The strategy helps respond to some of the big challenges that prominent scholars in the field are asking
Chapter Eight: Discussion and Conclusions

the reform community to face up to. One challenge is a matter of balance. Michael Fullan acknowledged in a critical discussion about the learning process models in this thesis at the American Education Research Association conference in Montreal that getting the right balance between vertically-driven improvement levers and horizontally-driven levers was one of the next big challenges in school improvement (M. Fullan, personal communication, April 12, 2005). Learning connections are a useful way of working out what the right balance is. Another challenge is a matter of depth. Cynthia Coburn issued the challenge a few years ago (Coburn, 2003) and participants in the field appear to be struggling to find out how to meet it. Learning connections both within and across the various interventions might help with that struggle.

This part of the theory about cross-initiative connections does not promote the idea of creating a highly organised matrix of learning connections. Research investigating complex partnerships within schooling improvement initiatives in the late 1990s indicates that some sort of master plan to create an orderly school improvement landscape is overly optimistic and may be counter-productive,

The difference in the partners’ theories of action and their approaches to some of the most pervasive dilemmas of schooling are deep and cannot be ignored. Even in a case where there was considerable agreement on goals and mission — when differences in approach were viewed as simply a matter of emphasis and not direct disagreement, where good relationships existed at the highest levels, and when significant funding was provided — different approaches to three of the basic dilemmas of schooling made it extremely difficult to make decisions and to carry out the collaborative work that school improvement required. (Hatch, 1998, p.24)

I believe that trying to create a tidy matrix of connections between various school improvement initiatives is an academic exercise which is likely to hinder rather than help the acceleration process. What is realistic is encouraging connections between initiative leaders in order to solve common problems. If anything, that strategy may appear messier than ever, at least for as long as it takes to get used to a broader

41 The three basic dilemmas of schooling that Hatch (1998) is referring to are, (i) how to establish wide support and foster innovation at the same time (p.14), (ii) how to balance the needs and interests of students, teachers, and society in the curriculum (p.17), and (iii) how to balance the need for autonomy with the benefits of support and direction (p.21).
interactive approach. But if that helps accelerate the learning process, then messiness may be an ingredient with which we must learn to live.

8.4. Implications

The theoretical end point of the thesis about consistent and connected learning focused on instructional improvement has some conceptual implications for the field of school improvement and some practical considerations for efforts in New Zealand.

8.4.1. Conceptual implications for the field of schooling improvement

Probably the most important conceptual implication of my theory for the field is to underpin our talk, written stories and actions with strong evidence of effectiveness. It sounds like such a simple task. It is just a matter of researchers helping school reform groups make links between the interventions that they develop and implement and the achievement gains that students are making in the classrooms. Yet there are precious few initiatives that are making those links explicit. Two possible conclusions that could be drawn from this situation are that most participants in the field either value other parts of their work ahead of making the links or that the task of making the links is much more complex than it looks.

In my view, it is the latter conclusion. One only has to read the findings of expert programme evaluators to verify that view. In relation to New York District #2, for instance, a team of researchers tried to test the assumption that teacher professional development is causally linked to all students meeting high academic standards (Harwell et al., 2000). Their efforts were inconclusive: “a multilevel data-analytic model is used to explore the effects of teacher engagement in professional development on student achievement. The results of these analyses reveal little influence of professional development on achievement in either mathematics or reading in District #2. However, due to various data collection and sampling difficulties, we view these findings as inconclusive” (p.1). If researchers from credible institutions such as the Learning Research and Development Centre at the University of Pittsburgh are struggling to make the links, it stands to reason that the task is a struggle for the masses of other researchers and reform groups operating in developed countries around the world.
Chapter Eight: Discussion and Conclusions

Rather than despair about this situation, it seems sensible for those in the field who are struggling to make the links to follow the lead of those precious few who are most advanced in doing that task. Part II of this thesis provided explanations of five international and two national examples of initiatives that appear to be the most advanced in this regard. So there are examples to follow and pitfalls to avoid. There is also an analytical and critically challenging language developed to help enact those examples. Two colleagues and I named this ‘learning talk’ (Annan et al., 2003) while another researcher has referred to it as ‘cognitive conflict’ (De Lima, 2001). It is the type of talk that encourages, rather than quells, collegial debates that: Brown and colleagues had with officials in England over the value of numeracy strategies (Brown, 1998; Brown et al., 2000); that Cook and associates had with Comer over social versus academic goals in the School Development Programme in the United States (Comer & Haynes, 1999; Cook et al., 1999); and that Harker and Nash had with Timperley and McNaughton and their colleagues about the quality of the evidence attached to some of the SEMO interventions (Harker, 2003; McNaughton, 2004; Nash, 2003; Timperley, 2004b). The health of the schooling improvement movement does not rest so much on who won those debates. Rather, it rests on the extent to which a broad spectrum of participants involved in schooling improvement initiatives learn from those debates and from their own evidence-informed debates and then use what they have learned to act more effectively. My main point here is that analysis, critique and challenge should be embraced in our everyday schooling improvement tasks and that there is little excuse nowadays for data-free talk, writing or action.

A second and complementary conceptual implication of my theory for the field of schooling improvement is that we are not going to accelerate things by continuing to use only those solutions that fit comfortably into our existing educational and related political and social systems. If we are serious about acceleration, we need to become courageous enough to try solutions that have proven to work to raise disadvantaged students’ academic achievement levels elsewhere. That means setting aside agendas to keep adult groups in the education system, such as teachers or politicians, satisfied and searching for effective national and international solutions that can best help solve local achievement problems. In New Zealand, that may mean considering more vertically-driven standardised ways of doing things and overseas it may mean
allowing horizontal reform networks to form that are not tightly constrained vertically.

So the 'professional' part of 'professional learning' is a matter of looking for the right solution to solve particular achievement problems, whether that means using a tightly scripted approach or a more liberal communal approach or something else. It's a matter of fit for purpose. Consequently, I do not believe professionalism and standardisation, or prescription (Fullan, 2005b), have to be portrayed as existing at two ends of a development continuum in schooling improvement. Rather, standardisation is one solution among many that professionals in the field should be considering to solve the achievement problems. As it happened, a strong element of standardisation was found in all seven initiatives in the investigations in Part III of this thesis. That might suggest that a characteristic of effective professionals working in the field of schooling improvement is that they are prepared to forfeit whatever degrees of autonomy they have to pursue a solution that works best.

8.4.2. Practical considerations for New Zealand’s school improvement movement

There is little doubt that the quality and quantity of outcomes-focused evaluations for schooling improvement initiatives in New Zealand needs immediate attention. At this point in time, numerous professional groups are working under the name of 'schooling improvement' without an accurate awareness of the impact that they are having on students' academic careers. Additionally, the centre, including government, the Ministry and the Education Review Office, have proven to be willing learning partners but they are trying to learn in a vacuum of context-specific achievement information. I have argued that this situation can be relatively easily rectified by altering the formative-summative balance in relation to assessing schooling improvement initiatives in favour of more summative evaluations. A re-balancing exercise of this nature might feel like a pendulum swing for professionals working in schooling improvement, and so it should, given the negligible amount of robust programme evaluation completed to date. Such a re-balancing exercise is not a pendulum swing towards some sort of national testing programme for all students in the country.

Despite New Zealand's reform movement being well behind that of England and the United States in terms of quality evidence of effectiveness, there is plenty of
evidence of a willingness both among initiative leaders and officials at the centre to address that situation. In line with New Zealand’s cultural norm of getting on with the job of fixing a problem once it has been identified, I have no doubt they will continue to ratchet up their efforts to address the current situation. However, I raised some reservations in Part II about continuing to accept a diverse range of evaluation methodologies and trying to replicate district-designed achievement information systems. Participants within individual initiatives have come around to thinking it is sensible to use common assessment tools across groups of schools. It seems highly likely that they would agree to scale up that endeavour to use common assessment measures across initiatives as long as the collaborations between the policy, research and development and practice communities remained committed to learning from one another. Additionally, the development of the achievement management system in Mangere is a tool from which many other localised reform groups could benefit. It would not only benefit other groups in terms of more efficient storage of their achievement information, it would also provide much richer comparison opportunities than are currently available. Again, such a scale up development would appear most palatable if it is motivated by strengthening local and central learning opportunities.

A final practical implication of the theory developed in this thesis for New Zealand is the opportunity to capitalise on the learning connections developed between policy developers, researcher-developers and practitioners and between vertically and horizontally-driven schooling improvement initiatives. Chapter Seven described policy-research-practice collaboration operating vertically at the top of the education system in the Numeracy Development Project and another one operating horizontally at the bottom of the system in the Mangere learning network. I argued that those sorts of collaborations developed in New Zealand mainly because the policy community is prepared to work and learn alongside the research and development and practice communities rather than setting itself apart from one or both of those communities. The latter happened in England and the United States and no such policy-research-practice collaborations were apparent in the five international initiatives. I believe that those two New Zealand-based policy-research-practice collaborations and others like them are a national treasure that can unlock the door to strong development and implementation networks across professional communities involved in schooling improvement.
A replication of the sorts of learning connections that occurred between the Numeracy Development Project and the district-level reform in Manurewa and between one of SEMO’s lead researchers (Timperley) and the national Literacy Professional Development Programme is likely to further strengthen those professional learning networks. So there is potential to create multiple communities of professionals collaborating within and across schooling improvement initiatives in New Zealand. Maximising the potential is not a matter of designing orderly connected diagrams to replace the messy reality of schooling improvement. It is more about equipping all reform groups with the sorts of inquiry-based learning skills and knowledge that the leaders of the Numeracy Development Project and the Mangere learning network have developed. The idea conjures up a generic professional learning course like the one called ERIC, which helped many teachers to better understand how to teach reading in the 1980’s. If such a generic course included the mastery learning techniques developed by the problem analyst in the Mangere learning network, reform groups will realise relatively quickly that it is better to search for the help of others to solve their problems than to struggle on alone and eventually stumble upon something useful.

8.5. Limitations of the thesis

There are three noteworthy limitations of this thesis. The first limitation relates to my ability to present an objective view about schooling improvement in New Zealand when I have been so closely involved in its development. I took considerable care to validate the accuracy of the New Zealand-based descriptions of events and to reduce the influence of my own biases. To this end, I carried out methodical searches of source documents and obtained participant validation checks. Documenting the SEMO project and the Mangere learning network posed the greatest threat to my impartiality. The review of the Numeracy Development Project was not as problematic because I have been well removed from the development and implementation of that initiative. It took considerable discipline to separate out my espoused theories based on personal biases from those formed through the empirical evidence which was implicit or explicit in the source documents and feedback from the policy developers. I methodically referenced the descriptive information and
spent considerable time and effort checking and re checking the accuracy of the information.

The second limitation is a methodological one. There was no comparison group for the seven successful initiatives. In other words, I did not find an equivalent sample of schooling improvement initiatives with little or no evidence of effectiveness to compare development and implementation characteristics. Consequently, there was no comparison point to validate the theoretical argument that the three common and interrelated characteristics, i.e. a focus on instructional practices, the use of standardised practices and learning connections, will help accelerate the learning process. In the absence of any evidence to clarify this point of comparison, the reader is left to assume that the less successful initiatives do not have the three interrelated characteristics. The point here is that they might exhibit those characteristics and if they do there must be other characteristics at play that distinguish the successful from the unsuccessful initiatives.

The third limitation is a philosophical one. The investigation was several steps removed from classroom activity. Yet Richard Elmore’s (2005) words of wisdom remind me that what happens in classrooms is at the heart of effective school improvement: “To succeed, school reform has to happen “from the inside out”” (p.3). Throughout my 30-year career in education I have held a belief that worthwhile educators are connected to classrooms and those that are not quickly get themselves out of touch. Due to my official status, I had to use research and evaluation documents and official reports as a proxy for a connection with classrooms. Although there were ample references in those documents about classroom activity, I admit it was definitely less than ideal. Looking at this limitation from a positive point of view, however, it did safeguard those working in and around classrooms from trying to relate to a researcher with official status. Additionally, it meant that one of the Ministry’s own was critically informing its involvement in schooling improvement.

8.6. Further Research

There are a number of research opportunities that can follow on from this thesis. At a pragmatic level, there are ample opportunities to help develop the evidence bases attached to individual school improvement initiatives in New Zealand. Additionally, the idea of having regular third-party summative evaluations provides scope for larger
scale comparison studies across initiatives than have been the case to date. Research of that nature would also create opportunities for comparative studies between the outcomes achieved in schooling improvement initiatives in New Zealand and those achieved through initiatives overseas. It would add another layer of analysis of the condition of the schooling improvement evidence base beyond what Borman et al (2003) have achieved to date. In New Zealand, our first priority is to produce a body of strong evidence so that we could participate in such an international comparison study.

At a conceptual level, it would be useful to find out about the characteristics of initiatives that have little or no evidence of effectiveness. Such studies might validate or invalidate the argument that the three common and interrelated characteristics found in the effective initiatives are worth replicating. I would also anticipate that the process of conducting a comparative research project of that nature would draw out competing theories about the definition of schooling improvement and which characteristics are valued by whom and why. So a part of such a study into less effective initiatives or a subsequent study again is to help participants resolve their differences in such a way that their espoused theories about effective school improvement are backed up with evidence.

8.7. Summary and conclusions

The overall problem of underachievement among disadvantaged students is one of the big challenges facing developed countries’ professional policy developers, researcher-developers and practitioners as well as the students themselves and their families. Learning how to solve that problem has proven to be incredibly difficult for all these groups. This thesis concentrated on the processes set up for professionals to learn effective ways of solving the overall problem in seven successful initiatives in order to help accelerate the learning process. A critical finding was the value of learning connections among policy developers, researcher-developers and practitioners to transform useful theoretical and practical ideas into practice. However, the evidence accumulated through the series of investigations in this thesis suggested learning connections on their own are insufficient to accelerate the learning processes. Equally important to get things moving is a sharp focus within the learning connections on instructional improvement and the development and implementation
of evidence-informed standardised practices. Hence, a theory for consistent and connected learning focused on instructional improvement.

I argue that spreading that theory from the few successful initiatives among the many initiatives struggling to get any traction would be a good start in accelerating professional learning of effective reform practices. However, a replication programme alone is not going to get the best out of professionals across the system. There is another more sophisticated level of development which can help accelerate learning among professionals. It is learning connections between improvement initiatives operating within the vertical and horizontal learning dimensions. Interconnected learning is promoted over learning through individual school improvement initiatives. New Zealand is well placed to capitalise on this development as it already has a good balance between vertical and horizontal learning levers and there are some useful interactions occurring across those two dimensions.

To conclude, the comparative analysis indicated that three conditions for change above others are useful for accelerating the learning of effective reform practices. The three conditions are a sharp focus on instructional improvement, consistent implementation of standardised practices and learning connections within and across initiatives. New Zealand's efforts stood up well in the comparative analysis in that its most advanced initiatives are leading the way in developing learning collaborations among the policy, research and development and practice communities. Efforts to create this interconnected approach and a commitment among senior policy developers to help develop stronger evidence of effectiveness, indicate to me that New Zealand's schooling improvement movement, at least at the most advanced end of developments, is moving on from the do-it-yourself No.8 wire tendencies of the past.
APPENDIX A
Letter for Policy Developers to Comment on Draft Information

Brian Annan
PhD Candidate
The University of Auckland

[Insert name]
[Insert position]
Ministry of Education
Pipitea Street
Wellington

Dear [Insert name],

Thank you for agreeing to comment on the two chapters of my PhD thesis, both of which are attached in this email. Please note that the two chapters and this letter have also been sent to you in hard copy in the mail.

My request to you is that you read the two chapters and feedback any comment you may have on the accuracy of the information. Please feel free to use your preferred method of comment. For instance, you could comment directly on the paper copies or create track changes in the electronic copies.

There is a stamped addressed envelop in your mail package if you choose to comment on the hard copy.

Your comment would be appreciated by Tuesday 1 November 2005.

Attached is a draft abstract which provides you with an overview of the thesis. Please feel free to make contact if you require any further information or if you wish to discuss any issues in relation to commenting on the two chapters.

Once again, thank you for agreeing to comment and I look forward to your response.

Regards

Brian Annan
APPENDIX B

Descriptions of Five International and Two National School Improvement Initiatives

England’s National Literacy and Numeracy Initiatives

<table>
<thead>
<tr>
<th>Developer</th>
<th>A lead bureaucrat, Michael Barber, with two professional teams established in national literacy and numeracy centres.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and Scale</td>
<td>National reform into all 22,000 primary schools across England (1997-2002)</td>
</tr>
<tr>
<td>The problem</td>
<td>Long-term neglect of disadvantaged students’ entitlement to quality schooling and the consequential academic underachievement problem.</td>
</tr>
<tr>
<td>Design Theory</td>
<td>Align curriculum and assessment policies and put schools under pressure to comply with national mandates.</td>
</tr>
<tr>
<td>Timeframe for Significantly Improving Student Achievement</td>
<td>3 Years</td>
</tr>
</tbody>
</table>
| Practices | • Scaled-up trials of literacy and numeracy teaching and national testing.  
• Developed standardised frameworks for literacy and numeracy, training packages, electronic information and teaching resources.  
• Mandated whole class teaching of one-hour scripted lessons in literacy and numeracy.  
• Conducted regular reviews of the students’ achievement levels against national standards through national testing.  
• Publicly disclosed schools’ achievement results. |
| Support Infrastructure | • National literacy and numeracy centres with directors.  
• An international research team providing formative support to the national centres.  
• A unified bureaucracy at the national, regional and local levels enabled Local Education Authority (LEA) regional directors to hand down the national mandates to line managers and advisors who, in turn, worked with contracted consultants to teach teachers and principals the standardised practices and to monitor their performance.  
• Principals were able to discuss implementation issues related to the reforms through collegial forums set up by LEA line-managers.  
• Teachers with expertise in the reform practices were selected as lead teachers to share their expertise with colleagues. |
Direct Instruction.

<table>
<thead>
<tr>
<th>Developer</th>
<th>Siegfried Engelmann, a researcher-developer at the University of Oregon.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and Scale</td>
<td>Comprehensive School Reform in approximately 300 U.S. schools (current).</td>
</tr>
<tr>
<td>The problem</td>
<td>Ineffective teaching causing many students to fail in school.</td>
</tr>
<tr>
<td>Design Theory</td>
<td>Student learning can be accelerated if teaching builds knowledge in</td>
</tr>
<tr>
<td></td>
<td>manageable bits then apply what is learnt to increasingly more challenging</td>
</tr>
<tr>
<td></td>
<td>learning tasks. This accelerated learning theory was made explicit in the</td>
</tr>
<tr>
<td></td>
<td>programme documentation.</td>
</tr>
<tr>
<td>Timeframe for</td>
<td>2-3 Years</td>
</tr>
<tr>
<td>Significantly Improving</td>
<td></td>
</tr>
<tr>
<td>Student Achievement</td>
<td></td>
</tr>
<tr>
<td>Practices</td>
<td>Highly scripted interactive reading, language and mathematics lessons</td>
</tr>
<tr>
<td></td>
<td>(science and social studies are integrated into the reading and language</td>
</tr>
<tr>
<td></td>
<td>lessons) to small groups of grade 1 and 2 students.</td>
</tr>
<tr>
<td></td>
<td>Whole class instruction beyond grade 2.</td>
</tr>
<tr>
<td></td>
<td>The use of standardised and field-tested curriculum resources.</td>
</tr>
<tr>
<td></td>
<td>A corrective reading programme for grade 4 students seriously behind</td>
</tr>
<tr>
<td></td>
<td>grade level in reading.</td>
</tr>
<tr>
<td></td>
<td>Regular reviews of achievement information through weekly conference</td>
</tr>
<tr>
<td></td>
<td>call.</td>
</tr>
<tr>
<td>Support infrastructure</td>
<td>A non-profit organisation called The National Institute of Direct</td>
</tr>
<tr>
<td></td>
<td>Instruction where Engelmann and a group of senior staff developed the</td>
</tr>
<tr>
<td></td>
<td>programme and monitored the impact of its implementation on student</td>
</tr>
<tr>
<td></td>
<td>achievement.</td>
</tr>
<tr>
<td></td>
<td>Institute consultants helped set up a management team in each school</td>
</tr>
<tr>
<td></td>
<td>then oversaw progress through a weekly conference call with the team</td>
</tr>
<tr>
<td></td>
<td>as well as weekly analytical reports.</td>
</tr>
<tr>
<td></td>
<td>A school management team with specific support functions:</td>
</tr>
<tr>
<td></td>
<td>o Principals as key change agents had to learn to be strategic and</td>
</tr>
<tr>
<td></td>
<td>instructional leaders. They were expected to develop a three-to-five</td>
</tr>
<tr>
<td></td>
<td>year strategic plan that detailed implementation procedures and</td>
</tr>
<tr>
<td></td>
<td>streamlined administrative systems so accountabilities and directives</td>
</tr>
<tr>
<td></td>
<td>were focused on curriculum and student performance;</td>
</tr>
<tr>
<td></td>
<td>o A full-time co-ordinator established a tight implementation and review</td>
</tr>
<tr>
<td></td>
<td>process. The co-ordinator helped teachers and principals learn the</td>
</tr>
<tr>
<td></td>
<td>prescribed practices and systems by delivering training sessions,</td>
</tr>
<tr>
<td></td>
<td>monitoring classroom lessons, providing formative feedback and</td>
</tr>
<tr>
<td></td>
<td>collating teachers' progress charts and achievement information for the</td>
</tr>
<tr>
<td></td>
<td>weekly conference call;</td>
</tr>
<tr>
<td></td>
<td>o Effective teachers were selected as peer coaches to help their</td>
</tr>
<tr>
<td></td>
<td>teacher colleagues to deal with implementation issues and to understand</td>
</tr>
<tr>
<td></td>
<td>the programme in depth.</td>
</tr>
</tbody>
</table>

Success For All.

<table>
<thead>
<tr>
<th>Developer</th>
<th>Robert Slavin and Nancy Madden, two researcher-developers from John Hopkins University.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and Scale</td>
<td>Comprehensive School Reform in over 2000 U.S. schools (current).</td>
</tr>
<tr>
<td>The problem</td>
<td>Existing schooling systems and practices were not working for disadvantaged students, particularly remedial programmes.</td>
</tr>
<tr>
<td>Design Theory</td>
<td>Teaching mastery skills and using collaborative learning techniques will improve student achievement</td>
</tr>
<tr>
<td>Timeframe for Significantly Improving Student Achievement</td>
<td>2-3 Years</td>
</tr>
<tr>
<td>Practices</td>
<td>• Grouping students into classes with similar reading levels.</td>
</tr>
<tr>
<td></td>
<td>• Teachers using proven teaching resources and practices in reading.</td>
</tr>
<tr>
<td></td>
<td>• Teachers engaged in eight-weekly assessments.</td>
</tr>
<tr>
<td></td>
<td>• Operating a family support team.</td>
</tr>
<tr>
<td>Support Infrastructure</td>
<td>• A non-profit organisation set up in Baltimore where Slavin, Madden and associates oversee the development, implementation and evaluation of the programme. With the headquarters in Baltimore, the directors can work closely with researchers from John Hopkins University.</td>
</tr>
<tr>
<td></td>
<td>• Consultants employed by headquarters work with in-school facilitators employed by the schools to introduce the programme to teachers and principals through a three-day training course. The consultants then step back to monitor progress through regular evaluative reports from the in-school facilitators and three in-depth site visits annually.</td>
</tr>
<tr>
<td></td>
<td>• Meanwhile the facilitators remain full time in the schools to provide on-going training and to set up and monitor instructional and system support strategies, such as a staff-support team and coaching partnerships.</td>
</tr>
<tr>
<td>References</td>
<td>(Borman &amp; Hewes, 2002; Borman et al., 2005; Northwest Regional Educational Laboratory, 2006; Rowan &amp; Correnti, 2006; Slavin &amp; Fashola, 1998; Slavin et al., 1992; Success For All, 2003)</td>
</tr>
</tbody>
</table>
### The School Development Programme

<table>
<thead>
<tr>
<th>Developer</th>
<th>Dr James Comer, a professor of child psychology at Yale University.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and Scale</td>
<td>Comprehensive School Reform in thousands of U.S. schools, the actual number is unspecified in available literature.</td>
</tr>
<tr>
<td>The problem</td>
<td>A main cause of underachievement in disadvantaged communities was homes and communities cause students to have development delays.</td>
</tr>
<tr>
<td>Design Theory</td>
<td>The design theory is based on a holistic assessment of the students’ developmental needs at entry to school. The idea is to find areas of developmental delay and create opportunities to accelerate the development process in those areas.</td>
</tr>
<tr>
<td>Timeframe for Significantly Improving Student Achievement</td>
<td>3-4 Years.</td>
</tr>
</tbody>
</table>
| Practices |  • Three school-based teams (school planning and management team, a student support team and a parent team) concentrate on improving school climate dimensions, such as raising teacher expectations of student performance, caring student-teacher relationships and creating fair and equitable discipline (There is no prescribed script for teachers to follow in the classroom).  
  • Key reform tasks are the development of a comprehensive school plan, staff development related to the plan and regular assessments of the teams’ joint impact on student learning.  
  • Standard rules that the teams are expected to adhere to are consensus decision-making, collaboration and no blame. |
| Support Infrastructure |  • A non-profit organisation called the Yale Child Study Centre at Yale University where Comer and a team of directors oversee programme development and monitor its success through several units, such as the Child and Adolescent Development Unit and the Programme Evaluation Unit.  
  • A national database of review and research information used to inform the development of the programme, support its implementation and evaluate its success.  
  • Implementation co-ordinators working from the Department of Education’s district offices help set up and monitor the school-based teams and liaise with the Yale Centre. |
New York District # 2.

<table>
<thead>
<tr>
<th>Developer</th>
<th>An experienced superintendent named Tony Alvarado with a small team of officials from the Department of Education’s district office.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and Scale</td>
<td>District-wide reform involving 32 publicly-funded schools (1987-1998).</td>
</tr>
<tr>
<td>The problem</td>
<td>Expected national achievement standards associated with the state and city-wide tests were too idealistic for schools serving disadvantaged communities.</td>
</tr>
<tr>
<td>Design Theory</td>
<td>Set achievable educational standards and develop and deliver tailored school improvement plans given the local constraints on individual schools.</td>
</tr>
<tr>
<td>Timeframe for Significantly Improving Student Achievement:</td>
<td>Gradual improvement over 10 years</td>
</tr>
</tbody>
</table>
| Practices | • District officials negotiated with each principal a school improvement plan with tailored standards, systems and practices (No one prescribed script for teachers).  
• Monitoring of each school’s impact on student achievement through Department officials regularly conducting on-site visits and analyses of city-wide test information.  
• The replacement of principals and teachers not delivering up to expectations (two-thirds of the principals and over half the teachers in District #2 were replaced from 1988-1998).  
• Interschool visits with a specific pedagogical purpose in mind for teachers to observe each other teaching. |
| Support Infrastructure | • The Department of Education district office team with strong leadership from Superintendent Alvarado  
• Contracted research to inform the district office team and school leaders. Research foci included school variability, continuous improvement, instructional leadership and assessing the impact of the professional development strand on student achievement.  
• Contracted educational consultants supported:  
  o principals with a formal induction programme to fulfil their responsibilities and mentor support, with particular attention paid to developing the improvement plans and appointing effective teachers;  
  o highly effective classroom teachers as in-school staff developers; and.  
  o teachers with pedagogical content knowledge and classroom management strategies. |
The Numeracy Development Project.

<table>
<thead>
<tr>
<th>Developer</th>
<th>A numeracy reference group since 2002 of policy developers, researchers, developers, publishers and lead practitioners (The reference group originated from a Maths and Science Taskforce in 1997 which became a Think Tank from 2000 to 2001).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and scale</td>
<td>A national professional development programme for all teachers of Year 1-8 students.</td>
</tr>
<tr>
<td>The problem</td>
<td>New Zealand students overall performed poorly in the 1995 Third International Mathematics and Science Study (TIMSS) in number, measurement, and algebra concepts.</td>
</tr>
<tr>
<td>Design theory</td>
<td>A comprehensive alignment of policies, practices and resources will help teachers understand number in depth and know about their students’ understandings of number which will enable them to help students use a range of strategies to solve mathematical problems.</td>
</tr>
<tr>
<td>Timeframe for significantly improving student achievement</td>
<td>5 years.</td>
</tr>
</tbody>
</table>
| Practices | • Developed a number framework which outlines typical student progress in number knowledge.  
  • A professional development programme for groups of teachers (Years 1-3, Years 4-6, Years 7-8 and Years 9-10) to learn how to  
    o conduct a diagnostic interview using a standard tool (Number Project Assessment — NUMPA) with each student to understand their number knowledge and strategy levels at the beginning and end of each year;  
    o plan knowledge lessons, strategy lessons and knowledge/strategy lessons;  
    o deliver lessons with more questioning and explanations of mathematical thinking; and  
    o conduct regular diagnostic interviews during the year using a second standard tool (Global Strategy Stage — GLOSS) to monitor students’ gains in understanding and use of number knowledge and strategies. |
| Support infrastructure | • A reference group to design and develop the project (The reference group originated from a Maths and Science Taskforce in 1997 which became a Think Tank from 2000 to 2001).  
  • A formative research team studying key aspects of the project (the number framework, the professional development programmes, the pedagogy being promoted and the achievement outcomes for the students).  
  • Facilitators led the professional development programme through seminars, classroom observations, modelling lessons followed by reflective discussion and talking with practitioners about content knowledge and pedagogical content knowledge consistent with the |
number framework.

- A comprehensive published programme both web-based and in print.


The Learning Network in Mangere.

<table>
<thead>
<tr>
<th>Developer</th>
<th>Co-constructed by lead-teachers and principals of the participating schools, researcher-developers and Ministry officials.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type and scale</td>
<td>A matrix of learning networks involving 8-16 schools in the Mangere district</td>
</tr>
<tr>
<td>The problem</td>
<td>What the school leaders were doing as well as what the researchers, developers and Ministry officials were doing to help the school leaders was not solving literacy the achievement problems.</td>
</tr>
<tr>
<td>Design theory</td>
<td>Strong learning connections between school leaders, researcher-developers and Ministry officials were the best way to work out how to solve the under achievement problems.</td>
</tr>
<tr>
<td>Timeframe for significantly improving student achievement</td>
<td>Seven Years.</td>
</tr>
</tbody>
</table>
| Practices | Participating schools agreed to use four inquiry-based learning practices to improve students’ ability to read and comprehend what they read. They agreed to:  
- use common assessment tools;  
- analyse the achievement information to identify achievement problems — reading comprehension was a priority problem;  
- adjust their teaching practices to address the identified achievement problems; and,  
- check the impact of the teaching on the problem through follow-up assessments. |
| Support infrastructure | - School leaders, researcher-developers and Ministry officials formed a learning network which had two components;  
  o an overarching policy-research-practice collaboration, which had three main functions; (i) it dealt with administrative issues (ii) it worked out that the domain knowledge worth spreading among all practitioners in the district was the analysis and use of student achievement data and (iii) it developed reform tools to help spread the domain knowledge;  
  o several smaller research-practice collaborations designed, implemented and evaluated classroom interventions.  
- A research team from The University of Auckland investigated the effective spread across the district of the domain knowledge in analysing and using student achievement data. |
Appendix B

- Additional funding and support from the Ministry’s monitoring and support division

References


References


References


Department of Education. (1987). New Zealand Department of Education Register of Progress and Achievement: Department of Education.


References


References


 References


References


Manurewa Enhancement Initiative Steering Committee, James Cook High School.

Ministry of Education.

*New Zealand Journal of Educational Studies, 39*(1), 153-156.

effective teaching of comprehension in culturally and linguistically diverse

Schools: 1993 to 2000 Student Performance: Course Taking, Test Scores, and

Ministry of Education. (1997a). *Proposal for Improving Schooling in Mangere and

Ministry of Education. (1997b). *Submission: Schooling In Mangere And Otara: A
Strategy For Improvement*. Wellington: Ministry of Education.

Ministry Project Office, Manukau City.

Gazette, 78*, 3-4.

Ministry of Education. (2000). *Submission (S99/1696): The Analysis and Use of
Student Achievement Data*. Wellington: Ministry of Education.

Wellington: Ministry of Education.
References


Ministry of Education. (2002b). *Submission (1026/03/0425/5): Schooling Improvement in Mangere and Otara 2002 - 2006 - Mangere AUSAD Management Team and the Otara Boards' Forum (Through Bairds Mainfreight Primary (1218) and Southern Cross Campus (452)*. Wellington: Ministry of Education.


Ministry of Education. (2004e). *SOAR @ Paeroa STAR Reading Results for Cluster November 2004*. Hamilton: Ministry of Education.

References


Ministry of Education. (2005f). *Submission (IO26/75/33/5) Schooling Improvement: PEN (Huntly- Ngaruawahia) approval of funding to implement the 2005 work programme*. Hamilton: Ministry of Education.


References


References


New Zealand Government. (1999). Education Act Amendment. 64A.


http://www.oecd.org/document/18/0,2340,en_2649_201185_2068050_1_1_1_1,00.html


Robinson, V. (2000). Reducing differential educational attainment: have we really tried. *Improving Schools, 3*(1), 40-44.


References


References


References


No Child Left Behind, Stat 1425 (2002).


