

<http://researchcommons.waikato.ac.nz/>

## **Research Commons at the University of Waikato**

### **Copyright Statement:**

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

The 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 the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

**Remapping the Assessment Landscape:  
Primary teachers reconstructing assessment in  
self-managing schools**

**A thesis submitted in fulfilment of the requirements of  
Doctor of Philosophy**

**at**

**The University of Waikato**

**by**

**Mary Frances Hill**

**2000**

**The University of Waikato**

**Hamilton**

**New Zealand**

## ABSTRACT

This participative case study took as its starting point a current “problem”, namely teachers’ concerns about assessment in self-managing primary schools. Prompted by teachers’ stress and confusion about their assessment role, I investigated how, in the 1990s, New Zealand teachers responded to multiple competing discourses of assessment, and the effects these discourses had on their teaching practices. Central to this thesis is the notion that, since compulsory schooling began in 1877, teachers’ assessment practices have been both constituted by and constitutive of dominant educational and assessment discourses. Taking account of the Foucauldian stress on discontinuity, this thesis argues that traces of teachers’ assessment practices can be found in earlier dominant discourses.

Evidence is presented to show that, in the latter half of the twentieth century, there was a discursive shift from education as a human right to education as a government investment, and that this shift refocussed public attention on both national standards and individual achievement. Through an examination of the structural changes to New Zealand education in the late 1980s and early 1990s, and the managerial discourses that accompanied them, the reasons why new technologies for measuring and monitoring learning and performance evolved are examined. But, in contrast to these accountability discourses, it is further argued that other professional assessment discourses were also competing for teachers’ attention at the time this investigation took place. A review of the international assessment research literature reveals evidence to support the view that, instead of increasing accountability mechanisms to raise standards, formative assessment strategies are critical in promoting learning and raising standards.

With these findings in mind, 12 teachers at two Waikato primary schools and 20 teachers at 13 other New Zealand primary and intermediate schools participated in semi-structured interviews about their assessment practices. Three teachers were then each observed in their classroom environment for a week. Foucault’s

analytics of power (namely discourse, disciplinary power, hierarchical observation, normalisation and the “examination”), are used to explain how and why, even though they valued formative assessment and child-centred pedagogies highly, many teachers predominantly employed summative assessment for accountability reasons. Due to increased managerial surveillance through the mechanisms of school and syndicate teamwork, appraisal, performance management and external audit, teachers have put themselves under pressure to account for their own practice. They have done this by finding ways to meet the documentation demands of the system, and in the process produced a summative shift in their assessment practice. Furthermore, although teachers were observed using formative assessment strategies known to improve learning, most did not identify these strategies as assessment and often understood the use of continuous summative assessment as formative. To address these findings, a case is made for professional development in assessment for school leaders and education officials as well as teachers. It is recommended that national and school policies of assessment, teacher appraisal and performance management be examined and revised to ensure that they produce predominantly formative rather than summative assessment practices.



## ACKNOWLEDGEMENTS

I am indebted to a large number of people whose support, assistance and encouragement made this thesis possible. Thanks to all who took an interest and urged me forward to complete this work. In particular, I want to extend my sincere gratitude to:

My supervisors, Dr Ian Calder, Professor Sue Middleton and Professor Ted Glynn, without whose high standards, challenging ideas, guidance, support and constructive criticism this work would not have been possible. In particular, Ian continually reminded me to return to the assessment literature and be sure I knew the field. Sue introduced me to Monsieur Foucault, his “toolkit” and his ideas, which allowed me to look through different lenses at the classroom data. And Ted’s meticulous reading assisted me in finding what was missing from the text and making it visible. Thank you all for your inspiration and enthusiasm.

The staff and children in schools throughout New Zealand who participated in this study and so willingly shared their time and ideas with me. Special thanks are extended to the three Waikato schools and especially the three teachers at them who opened their classrooms to me and gave so freely of their expertise. Rose, Luke and Anne – you know who you are, and to you I am forever grateful.

My friends and colleagues, both in New Zealand and overseas, who were instrumental in my undertaking this work and who encouraged me to complete it. These include Ministry of Education officials, Education Review Office members, principals and teachers, as well as many colleagues at Waikato and other universities, colleges of education and polytechnics. In particular, my sincere thanks are extended to everyone at the School of Education, University of Waikato, who assisted and supported me. Although too many to name, you have all been critical in my completing this work.

And all my close friends and family who have “seen me through”. We all knew it would be a long and arduous journey but your faith in me never wavered. Thanks Debbie for being my “test pilot”, and Pam and Clare for your personal care, concern and support throughout. Sincere thanks to Mark Bathurst for the meticulous proof reading. Thank you to my whole family for your interest and encouragement, particularly to my mother, Zoë, for seeing it finished; and last, but by no means least, to Mark who cared for me, gave me time and space to work, and had faith in my ability to complete this thesis.

# TABLE OF CONTENTS

|                                                                                                  |     |
|--------------------------------------------------------------------------------------------------|-----|
| <b>Abstract</b>                                                                                  | ii  |
| <b>Acknowledgements</b>                                                                          | iv  |
| <b>Table of Contents</b>                                                                         | vi  |
| <b>List of Tables</b>                                                                            | xi  |
| <b>List of Figures</b>                                                                           | xii |
| <br>                                                                                             |     |
| <b>Chapter One: Introduction</b>                                                                 | 1   |
| Mapping the Terrain                                                                              | 2   |
| My Place on the Educational Landscape                                                            | 3   |
| Using Foucault's "Toolkit"                                                                       | 5   |
| Disciplinary Power                                                                               | 6   |
| The "Examination"                                                                                | 8   |
| Discourse                                                                                        | 10  |
| The Structure of the Thesis                                                                      | 12  |
| <br>                                                                                             |     |
| <b>Chapter Two: Assessment Discourses in New Zealand Education</b>                               | 17  |
| Before Compulsory Schooling: 1840-1877                                                           | 19  |
| Testing Knowledge and Taking Places                                                              | 19  |
| The Case for Compulsory Schooling                                                                | 21  |
| Standards, Proficiency and Examinations: 1877-1937                                               | 22  |
| Examining Standards                                                                              | 23  |
| The Inspectors as the Public's Educational Watchdogs                                             | 24  |
| Measuring the Efficiency of Schools                                                              | 25  |
| Harbingers of Reform                                                                             | 28  |
| The Natural and the Normal: 1920s-1960s                                                          | 30  |
| The Natural                                                                                      | 31  |
| The Normal                                                                                       | 34  |
| Post-Proficiency Progressivism                                                                   | 36  |
| Accountability Discourses                                                                        | 40  |
| Assessment and Learning - the Rise of Criterion-referencing<br>and Individualisation: 1960s-1989 | 41  |
| Improving Teaching and Learning or<br>Maintaining Standards?                                     | 45  |
| Keeping School Records                                                                           | 49  |
| Summary and Comment                                                                              | 50  |

|                                                                                         |            |
|-----------------------------------------------------------------------------------------|------------|
| <b>Chapter Three: Self-managing Schools and Managerial</b>                              |            |
| <b>Assessment: 1984-1995</b>                                                            | <b>54</b>  |
| Economic Rationalism, Corporate Managerialism<br>and the Marketisation of Education     | 55         |
| Economic Rationalism, Managerialism and the<br>“Reform” of Primary Education            | 58         |
| <i>Administering for Excellence</i> (the Picot Report)<br>and <i>Tomorrow’s Schools</i> | 62         |
| <i>Tomorrow’s Standards</i>                                                             | 63         |
| National Standards and Monitoring                                                       |            |
| National Performance                                                                    | 65         |
| Assessing the Effectiveness of Schools                                                  | 66         |
| Assessment of Students                                                                  | 67         |
| Assessment and the Dual Cultural Heritage                                               | 68         |
| Pre-service and Inservice Training in<br>Assessment Techniques                          | 69         |
| The Achievement Initiative                                                              | 69         |
| The New Zealand Curriculum                                                              | 70         |
| The Education Review Office                                                             | 73         |
| Technocratic and Managerial Assessment                                                  | 75         |
| Summary and Comment                                                                     | 78         |
| <br><b>Chapter Four: Competing Discourses in Educational Assessment:</b>                |            |
| <b>The Research Literature</b>                                                          | <b>80</b>  |
| Educational Assessment Discourses                                                       | 81         |
| Formative Assessment                                                                    | 86         |
| Convergent Assessment                                                                   | 87         |
| Divergent Assessment                                                                    | 89         |
| Summative Assessment                                                                    | 93         |
| Interactions between Purposes                                                           | 98         |
| Competing Discourses of Assessment                                                      | 100        |
| Overseas Research into Teachers’ Assessment Practices                                   | 105        |
| Research on Teachers’ Assessment Practices in New Zealand<br>in the Early 1990s         | 120        |
| Summary and Comment                                                                     | 127        |
| <br><b>Chapter Five: Researching Teachers’ Practice</b>                                 | <b>130</b> |
| Multiple Discourses of Educational Research                                             | 131        |
| A Participatory Case Study Approach                                                     | 137        |
| Identifying the Research Issue and Questions                                            | 139        |
| Research Questions                                                                      | 140        |
| Determining Ethical Guidelines                                                          | 141        |
| Trustworthiness                                                                         | 141        |

|                                                  |     |
|--------------------------------------------------|-----|
| Dependability and Consistency                    | 146 |
| Predispositions of the Researcher                | 147 |
| Gaining Ethical Approval                         | 148 |
| Collecting and Storing Data                      | 151 |
| Interviews                                       | 153 |
| Observations                                     | 158 |
| School Records                                   | 160 |
| Generating and Testing the Analytical Statements | 160 |
| Interpreting and Illuminating the Findings       | 164 |
| Narrative Reporting                              | 165 |
| Timing                                           | 166 |
| Summary and Comment                              | 167 |

## **Chapter Six: Setting the Scene: Assessment Stories from**

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| <b>Two Schools</b>                                                | 169 |
| Stories from the Schools                                          | 170 |
| Arawhata School                                                   | 170 |
| Waimaka School                                                    | 180 |
| Discourses of Assessment                                          | 185 |
| Accountability Discourses                                         | 186 |
| Accountability to Governmental Requirements                       | 187 |
| Accountability to Colleagues                                      | 198 |
| Accountability to Students and Parents                            | 200 |
| Formative Discourses of Assessment                                | 202 |
| Informal Testing                                                  | 204 |
| Continuous Assessment                                             | 206 |
| Student Self-assessment                                           | 213 |
| Teachers at the Intersection of Multiple and Competing Discourses | 214 |
| Summary and Comment                                               | 217 |

## **Chapter Seven: Competing Discourses and Teachers' Assessment**

|                                                                                |     |
|--------------------------------------------------------------------------------|-----|
| <b>Practices</b>                                                               | 219 |
| Teachers' Responses to Multiple Competing Discourses of Assessment             | 220 |
| Three Approaches to Classroom Assessment                                       | 224 |
| Unit Assessment                                                                | 224 |
| Head Note Assessment                                                           | 225 |
| Integrated Systematic Assessment                                               | 226 |
| Discussion                                                                     | 227 |
| Delving Further into the Assessment Approaches at Arawhata and Waimaka Schools | 230 |
| Other Teachers, Other Schools                                                  | 233 |
| The Interview Questions                                                        | 237 |

|                                                                      |         |
|----------------------------------------------------------------------|---------|
| The Teachers' Responses                                              | 239     |
| The Effects on Teachers' Assessment Practices                        | 240     |
| Unit Assessment                                                      | 240     |
| Head Note Assessment                                                 | 247     |
| Integrated Systematic Assessment                                     | 253     |
| Other Approaches                                                     | 258     |
| Summary and Comment                                                  | 260     |
| <br><b>Chapter Eight: Assessment in Context: Three Portraits</b>     | <br>265 |
| Rose                                                                 | 268     |
| A Checking Approach to Assessment                                    | 269     |
| A Clash of Discourses                                                | 270     |
| The Need to Account: "I'd have a printer in my head"                 | 271     |
| Three Instruments of Disciplinary Power                              | 272     |
| Summative Assessments Recorded                                       | 275     |
| Assessment for Formative Purposes                                    | 279     |
| Making Assessment Manageable                                         | 282     |
| Competing Discourses of Assessment                                   | 286     |
| Luke                                                                 | 291     |
| "I'm a bit old fashioned": Using Tests for Assessment                | 292     |
| Planned Formative Assessment: Using Test                             |         |
| Information in Teaching                                              | 293     |
| Dot, Tick, Two Ticks: Summative Assessment Codes                     | 299     |
| Encouraging Evaluation                                               | 302     |
| On Becoming a Junior School Teacher                                  | 304     |
| Dynamic Assessment in Action                                         | 306     |
| Assessment or Evaluation: Competing Discourses?                      | 308     |
| Anne                                                                 | 311     |
| "A circular diagram in my mind: assess, plan, teach"                 | 312     |
| Using Records to Increase the Efficiency of Formative                |         |
| Assessment                                                           | 313     |
| Shared Teaching: Implications for Assessment                         | 322     |
| Theorising Anne's Practice                                           | 323     |
| Disciplining Learning – and Teaching                                 | 325     |
| Summary and Comment                                                  | 326     |
| <br><b>Chapter Nine: Lessons from the Classroom: Revisioning the</b> |         |
| <b>Assessment Landscape</b>                                          | 329     |
| A Summative Turn in Response to Accountability Demands               | 332     |
| Multiple Stories of Assessment                                       | 335     |
| Improving Learning through Formative Assessment                      | 338     |
| Implications for Professional Development                            | 340     |
| The Context of Self-managing Schools                                 | 342     |
| Concertive Control: School Policy and Teamwork                       |         |

|                                                                                                                                       |     |
|---------------------------------------------------------------------------------------------------------------------------------------|-----|
| as Normalisation Devices                                                                                                              | 344 |
| Making the “Examination” Formative at the School level                                                                                | 349 |
| The Impact of Research on Practice                                                                                                    | 352 |
| National Policies of Assessment and their Impact on<br>Assessment, Teaching and Learning                                              | 354 |
| Impacts on My Own Practice                                                                                                            | 357 |
| Limitations of the Study and Areas for Further Research                                                                               | 360 |
| Conclusion                                                                                                                            | 364 |
| <b>References</b>                                                                                                                     | 366 |
| <b>Appendices</b>                                                                                                                     | 395 |
| Appendix A:                                                                                                                           |     |
| Informed Consent Formats and Letters                                                                                                  | 395 |
| Appendix B:                                                                                                                           |     |
| Interview 1 Questions                                                                                                                 | 407 |
| Appendix C:                                                                                                                           |     |
| Interview 2 Questions:                                                                                                                | 408 |
| Appendix D:                                                                                                                           |     |
| Interview 3 Questions                                                                                                                 | 409 |
| Appendix E:                                                                                                                           |     |
| Teacher Quotations Pilot                                                                                                              | 418 |
| Appendix F:                                                                                                                           |     |
| Executive Summary of the Recording and Reporting<br>Action Research Project Carried Out for the Ministry<br>of Education in 1997-1998 | 421 |

# LIST OF TABLES

## **Table 7.1**

|                                                                                                |     |
|------------------------------------------------------------------------------------------------|-----|
| Overview of the schools from which the teachers were<br>drawn in the national interview sample | 235 |
|------------------------------------------------------------------------------------------------|-----|

## **Table 7.2**

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| Overview of the teachers age, gender, years of experience<br>and class taught in the national interview sample | 236 |
|----------------------------------------------------------------------------------------------------------------|-----|



# LIST OF FIGURES

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 6.1</b>                                                                                               |     |
| Funnel Model of School-wide assessment generated by the teachers at Arawhata School                             | 177 |
| <b>Figure 8.1</b>                                                                                               |     |
| Extract from Rose's roll book: Term 1, Speaking                                                                 | 277 |
| <b>Figure 8.2</b>                                                                                               |     |
| Reading records from Rose's roll book                                                                           | 278 |
| <b>Figure 8.3</b>                                                                                               |     |
| Rose's Checklist for her algebra unit                                                                           | 285 |
| <b>Figure 8.4</b>                                                                                               |     |
| Luke's basic facts record                                                                                       | 294 |
| <b>Figure 8.5</b>                                                                                               |     |
| Luke's progressive reading Level records                                                                        | 295 |
| <b>Figure 8.6</b>                                                                                               |     |
| P.A.T. results from Luke's roll book                                                                            | 296 |
| <b>Figure 8.7</b>                                                                                               |     |
| Luke's roll book page for recording progress against achievement objectives in the number strand in mathematics | 300 |
| <b>Figure 8.8</b>                                                                                               |     |
| Anne's data book page for recording achievement in PE                                                           | 316 |

**Figure 8.9**

Page from Anne's data book showing how she recorded  
achievement in oral language 317

**Figure 8.10**

Page from Anne's data book showing how she kept track of  
high frequency words in reading vocabulary 318

**Figure 8.11**

Page from Anne's data book showing how she kept track of  
achievement in concepts about print 319

**Figure 8.12**

Pages from Anne's data book showing how she recorded  
running records on a daily basis 320

# **CHAPTER ONE**

## **INTRODUCTION**

The last decade and a half have seen a discursive shift from education as a human right to education as a government investment. There have been many telling analyses of policy and programmes in these new times which suggest the dangers of these fundamentally economic ways of constituting the purposes and the consequences of education. However, the effects of managerial discourses in actual classrooms – preferred pedagogies, enacted curriculum, assessment practices, classroom talk and allocation of time – are rarely studied and difficult to ascertain. (Comber, 1997, p. 390)

I suspect most of us can understand Foucault's preference, namely to be already within discourse, to be "borne away beyond all possible beginnings", for beginning is never simple. (Smart, 1985, p. 10)

Between 1989 and 1995 (when this investigation began) the New Zealand education system was completely restructured. Initially the administration of education was changed to devolve responsibility and decision-making to each school and its local community. Following hard on the heels of this change came the reconstruction of the national curriculum as comprising a specific set of hierarchically ordered achievement objectives. It was left to schools and the teachers within them to design and implement assessment that meets what some commentators (for example, Nisbet, 1992) suggest are the irreconcilable demands of improving student learning and accountability to both the government and the local community. This thesis investigates how teachers in these self-managing primary schools have responded to the multiple competing discourses of assessment and the effects these discourses have had on their assessment and teaching practices.

Research into the assessment practices of primary school teachers in the past have tended to present a "view from the top" (Middleton & May, 1997) in that the historical material has mostly been drawn from Education Acts and Department and Ministry of Education policy, regulations, curriculum and guidelines documents (McCulloch, 1987;

Olssen, 1987). While acknowledging that such official points of view are critical to understanding educational phenomena such as assessment, other readings are both possible and necessary. D. McKenzie (1984) warned that historians have too often ignored not simply the classroom, but the dynamics of the teaching-learning process itself. My interest in this topic was sparked by the discontent of teachers as they struggled to implement assessment practices that “worked”. By “worked”, I mean that teachers had to come up with assessment techniques that met the expectations of their superiors (mainly the Education Review Office and principals), that were manageable in the classroom, and that gave the information that they needed to operate their day-to-day programme in the classroom. My joint roles of teacher educator and researcher served to make me aware that it was not enough to study this situation from outside the arena of practice. I felt compelled to join with some of these teachers as they experienced the process of working to improve their assessment practices. In assisting them, and in investigating their journey as they undertook it, I was able to document what appeared to me as a view from below, or perhaps, more colloquially, a view from the “chalkface”.

## **Mapping the Terrain**

The professional knowledge “landscape” is a metaphor used by Clandinin and Connelly (1995) that

allows us to talk about space, place and time. Furthermore, it has a sense of expansiveness and the possibility of being filled with diverse people, things and events in different relationships. Understanding professional knowledge as comprising a landscape calls for a notion of professional knowledge as composed of a wide variety of components and influenced by a wide variety of people, places and things. Because we see the professional knowledge landscape as composed of relationships among people, places and things, we see it as both an intellectual and a moral landscape. (pp. 4-5)

One’s place in this educational landscape (Clandinin & Connelly, 1995; Clandinin & Connelly, 1996) tends to alter the view of the environment. Viewed from the top, educational assessment in the 1990s may appear to be reasonably unproblematic. In this view, school management techniques that combine assessment data for

individuals are thought to assist teachers to see patterns of achievement and progress (or the lack of it). From this perspective, these patterns can be analysed for clues to improve achievement and design programmes that will be applied skilfully by teachers to address the barriers to learning that were identified through the analysis. In addition, the information gained in the process should be used to report achievement and progress to both the community and the school system.

### My Place on the Educational Landscape

Viewed from my place on the assessment landscape the picture appeared more complex. I had 20 years experience of teaching at primary and tertiary levels of education. Since the introduction of *Tomorrow's Schools* (Lange, 1988) I had been employed as a lecturer in education at the University of Waikato. In addition to teaching courses about learning, teaching and assessment to pre-service teachers, I had also been contracted by the Ministry of Education to provide professional development in assessment for teachers (Calder, 1996; Faire, Brown, & Yates, 1994)<sup>1</sup>. I had also undertaken a small action research project in one teacher's classroom in a local school that allowed me to participate as a staff member in a large urban school setting for up to a day a week during 1995 (M. Hill & Jones, 1995). From my interactions with teachers in this school and through my teaching and involvement in professional development courses, there appeared to be an assessment frenzy taking place. Rather than the supposed rational, orderly approach indicated by the top-down view, assessment at the chalkface appeared to be a contested area, a cause of stress and confusion. Although it appeared that the curriculum had been spelt out more clearly than ever before, and the regulations introduced in 1993 had seemingly clarified and specified in simple language what schools were required to do, schools and teachers were seeking help more than ever before. Why had this area of teaching become so problematic for schools and teachers? To investigate this problem, I participated in what Hultqvist (1998) regards as social

---

<sup>1</sup> Faire was my married name, which I used until 1995, when I reverted to using my own name, Hill.

scientific thinking, which “involves problematizing and diagnosing current problems, and finding ways to solve them” (p. 93).

Traditionally, social science research has attempted to provide an objective view of reality, independent of particular social and moral values and interests, in order to get at what seems to be the “truth” of the matter. In my Master of Education thesis, which was also an investigation of teaching and learning, I wrote within what has been termed a positivist frame of reference (Jones, 1992; Olssen, 1987). I had learnt that thesis writing demanded I distance myself from the phenomena I was studying. However, while I acted as the invisible, neutral, observing presence in my written text, I was aware of the way in which my framing of the research questions, my data-gathering techniques, my analysis of the data and my interpretation of the results framed and shaped what I wrote. I had read Kuhn’s work about the nature of paradigms – how what counts as scientific knowledge is shaped by its historical and social context – and postmodern and feminist critiques. In undertaking an investigation of assessment practices I wanted to acknowledge the relativity of truth, the historical and social embeddedness of teachers’ practices and the partiality of knowledge. I could see that the intentions of the policy makers were “to improve students’ learning and the quality of learning programmes” (Ministry of Education, 1993a, p. 24), but it was also evident that the policy and curriculum rhetoric were not, in and of themselves, enough. In addition, as well as producing an academic treatise, I wanted to help teachers build their confidence and improve their practice of assessment. At the same time, however, I was very well aware of the clash between the two worlds in which I was working in undertaking this thesis – the academic and the professional. The academic requirements of thesis writing required that my work be original, demonstrably my own work and theoretical.

A thesis is a sustained piece of original scholarship. It must show a thorough knowledge of, and engage critically with, and make an original contribution to, the literature in a discipline or field. (Centre for Graduate Studies School of Education, 2000, p. 3)

But in my other world I knew that what teachers needed was not more theory to implement. As one said

My own gripe is the impractical and irrelevant verbose screeds that issue from the bespectacled boffins in the universities/colleges of education that don't help me at the chalk face! (Primary School Principal)

Put a different way, “while receptiveness to critical theories grows in the academy, the estrangement of critical theory from its audience characterises its impact on policy and its resonance in the lives of people outside the academy” (Lather, 1991, p. xvii). I was concerned, therefore, that my description resonate with teachers’ practice. I needed a research design that was flexible enough to allow the collection of information about teachers’ beliefs and practices, while, at the same time, informing practice and advancing theory (Cohen & Manion, 1989; Elliott, 1991; Kemmis & McTaggart, 1988; Lather, 1986). To achieve this, I adopted the recursive strategy of recycling my developing theoretical interpretations to teachers for comment, elaboration and criticism.

### **Using Foucault’s “Toolkit”**

Further reading (for example, Ball, 1990; Middleton, 1990) introduced me to the work of Michel Foucault. The question Foucault asked in much of his work was not “what is true?” but “how is truth created?” and “what are the effects of socially constructed truth?” (Simola, Heikkinen, & Silvonen, 1998). Middleton suggests that Foucault’s methodology for inquiries “provides a useful ‘tool kit’ with which to think historically and geographically” (Middleton, 1988, p. 3). His writing, certainly, encouraged me towards a re-examination of the assessment literature, searching for different ways in which the truths about assessment had been created.

Foucault examines practices and techniques for the production of truth, the constitution of the truth-willing subject, and the separation of true and false; that is, techniques of discourse, self, and government. We call the ensemble of these techniques technologies of truth. (Simola et al., 1998, p. 64)

As well, I found critiques that made me very sceptical of even mentioning Foucault’s name. For example, Apple stated that a

Large part of what is called “critical educational studies” has tended to be all too trendy. It moves from theory to theory as

each new wave of elegant meta-theory (preferably French) finds its way here [the United States]. (Apple, 1993, p. 299)

However, continued reading, talking and thinking led me to the conclusion that there are ways in which a Foucauldian perspective, when used reflexively and appropriately (Apple, 1998), could help me to illuminate the workings of power in the classroom assessment context. For example, Foucault advised against going about social research from the perspective of a top-down view of “grand theory”. In contrast he favoured studying empirically the everyday workings of life in particular contexts “with power at its extremities, in its ultimate destinations, with those points where it becomes capillary, that is, in its more regional and local forms of institutions” (Foucault, 1980, p. 96). With the exception of a handful of scholars (for example, Bernstein, Bourdieu and Passeron, and others who have drawn on their work) educational researchers have paid little attention to the micro-level functioning of power in pedagogy that Foucault recommends (Gore, 1998). Using a micro-level analysis of classroom pedagogy I believed it might be possible to illuminate some of the policies and practices and their impact on teaching and learning that have led teachers to the assessment frenzy described earlier.

### Disciplinary Power

Elaborating on the invisibility and pervasiveness of power in modern society, Foucault argued that “the eighteenth century invented, so to speak, a synaptic regime of power, a regime of its exercise within the social body rather than from above it” (Foucault, 1980, p. 39). In his “middle period”, Foucault (for example, 1977, 1978, 1980, 1982) studied the mechanisms of power in institutions which characterise modern Western societies. There he explained disciplinary power by contrasting it with what he terms “sovereign power”. Power as sovereignty centres actors as the wielders of power as, for example, the King with power over life and death, or governing or ruling groups favoured in decision making. In this landscape, power is something that people own, and that ownership can be redistributed among groups to challenge inequities (Popkewitz & Brennan, 1998). Popkewitz and Brennan show how the concept of power as sovereignty



is found in much of the sociology of school knowledge, where it is used to explain the origins of domination and subjugation in society.

For example a sovereignty notion of power is embodied in current educational literature that sees social interests inscribed in reform reports and government policies that argue for “back to basics” curriculum. The consequences of the reforms, it is argued, is to reproduce gender, racial, and class distinctions in society. (p.17)

Because sovereign power is attached to actors who have legitimacy to make decisions and allocate values within communities, a central premise is that societies contain groups, social interests, and forces “that historically have formed and whose practices dominate and repress other groups” (Popkewitz & Brennan, 1998, p. 17)

In contrast, while acknowledging that sovereign power exists and exerts pressure, Foucault describes disciplinary power as circulating rather than being possessed, productive and not necessarily repressive, existing in action, functioning at the level of the body, often operating through “technologies of self” (Gore, 1998, p. 233). Foucault described citizens within a disciplined society as docile bodies. “A body is docile that may be subjected, used, transformed and improved” (Foucault, 1977, p. 136).

He argued that from the eighteenth century, rather than treating populations en masse through subjection and domination, new governmental techniques emerged that enabled an infinitesimal power over the active body, working it individually, exercising upon it a subtle coercion “of obtaining holds upon it at the level of the mechanism itself – movements, gestures, attitudes, rapidity” (1977, p. 137).

Then there was the object of control: it was not or no longer the signifying elements of behaviour or the language, but the economy, the efficiency of movements, their internal organization; constraint bears upon the forces rather than upon the signs... Lastly, there is the modality: it implies an uninterrupted, constant coercion, supervising the processes of activity rather than its result and it is exercised according to a codification that partitions as closely as possible time, space, movement. These methods, which made possible the meticulous control of the operations of the body, which assured the constant subjection of its forces and imposed upon them a relation of docility-utility, might be called “disciplines”. (Foucault, 1977, p. 137)

Foucault described the gradual emergence of disciplinary power as a “multiplicity of often minor processes, of different origin and scattered location which overlap, repeat, or imitate each other, support each other, or distinguish themselves from one another...converge and gradually produce the blueprint of a general method” (1977, p. 138). He describes these processes as being at work in secondary schools “at a very early date” (p. 138), then moving into primary schools, hospitals, and the military. Using the institution of the prison as an exemplar, Foucault provided a careful elaboration of how specific techniques of power were used to govern the penal population. Of critical importance to these techniques is power/knowledge. Foucault referred to its presence as “le regard”, meaning surveillance in the form of an uninterrupted, synoptic and individualising gaze (Meadmore, 1993). He used Bentham’s panopticon to epitomise the use of surveillance in a prison, where, through making convicts visible in back-lit cells, observable at any time by virtually anyone from the position of a central tower, each individual becomes a “case”.

In fact, any panoptic institution, even if it is rigorously closed as a penitentiary, may without difficulty be subjected to such irregular and constant inspections: and not only by the appointed inspectors, but also by the public; any member of society will have the right to come and see with his own eyes how the schools, hospitals, factories, prisons function. There is no risk, therefore, that the increase in power by the panoptic machine may degenerate into tyranny; the disciplinary mechanism will be democratically controlled, since it will be constantly accessible to “the great tribunal committee of the world”. (Foucault, 1977, p. 207)

### The “Examination”

Rather than the back-lit cells in Bentham’s panopticon, in schooling technical means are used in order that such disciplinary power may operate. These technical means include strategies such as examinations, tests, inspections, profiles, registers, classroom and playground duties and the like (Gore, 1998; Meadmore, 1993). In *Discipline and Punish*, Foucault set out how disciplinary power “makes” individuals through the use of “hierarchical observation, normalising judgement and their combination in a procedure that is specific to it, the examination” (1977, p. 170).

The examination combines the techniques of an observing hierarchy and those of normalizing judgement. It is a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish. It establishes over individuals a visibility through which one differentiates them and judges them. (Foucault, 1977, p. 184)

According to Foucault, the “examination” as a general concept has its own history which should be documented (Meadmore, 1995). In this study various forms of the examination are investigated: teacher assessment of student work, inspections of schools, records of progress, ranking children against the normal curve, and student self-assessment. All of these forms of assessment can be analysed from the point of view of the “examination” as a “technology of government” which pays attention to

actual mechanisms through which authorities of various sorts have sought to shape, normalize and instrumentalize the conduct, thought, decisions and aspirations of others ... to achieve the objectives they consider desirable. (Miller & Rose, 1990, p. 8)

In this “slender technique” (Foucault, 1977, p. 185) of the examination the techniques of an observing hierarchy and normalising judgement are combined. At the very heart of teaching lies the mechanism of hierarchized, continuous and functional surveillance by which disciplinary power becomes “an ‘integrated’ system, linked from the inside to the economy and to the aims of the mechanism in which it was practised” (Foucault, 1977, p. 176). Through such continuous surveillance, Foucault argues that disciplinary power is absolutely “indiscreet” “since it is everywhere and always alert, since by its very principle it leaves no zone of shade and constantly supervises the very individuals who are entrusted with the task of supervising; [but at the same time] absolutely ‘discreet’, for it functions permanently and in silence” (1977, p. 177).

In combination with hierarchical observation, normalising judgement that classifies all behaviour in the field between good and bad, enables teachers, and those that might supervise them, to make assessment decisions. Discipline, then, rewards by the application of awards, such as a tick, attaining a certificate or gaining a higher rank; but it also punishes by reversing this process, through the allocation of a lower grade or rank or even by exclusion from “the norm” and by classification as abnormal or other.

Being at the heart of the procedures of discipline, assessment, as the examination, “manifests the subjection of those who are perceived as objects and objectification of those who are subjected” (Foucault, 1977, pp. 184-5). Therefore, in times of rupture and political change, it stands to reason that the form assessment takes may be in dispute, a site of struggle and tension. At times such as these, when dominant discourses are challenged and sometimes overtaken by more powerful discourses, the very techniques that govern how subjects are made in society, such as the examination (assessment and documentation), will be transformed to support the more dominant discourse/s.

### Discourse

By discourse Foucault meant a form of power that circulates in the social field and can attach to strategies of domination as well as to those of resistance (Johannesson, 1998). It is not just what is said but the total context in which it is said and in which it is given meaning. In other words

Discourse is constituted by the difference between what one could say correctly at one period...and what is actually said. The discursive field is, at a specific moment, the law of this difference. (Foucault, 1978, cited in Johannesson, 1998, p. 305)

The word *discourse* is derived from the Latin *discurrere* – running to and fro, to run different ways – and, latterly, from Medieval Latin, *discursis*, meaning argument (HarperCollins, 1998). But from a Foucauldian perspective, the concept of discourse refers to the materiality of practices as much as it refers to the spoken and written word.

Discourses, in Foucault’s work, are ways of constituting knowledge, together with the social practices, forms of subjectivity and power relations which inhere in such knowledges and the relations between them. Discourses are more than ways of thinking and producing meaning. They constitute the “nature” of the body, unconscious and conscious mind and emotional life of the subjects which they seek to govern. (Weedon, 1987, p. 108)

This means that discursive practices include the silences in discourse as much as they do the words, and patterns of behaviour and the organization of structures of society are also parts of discourse.

Discourses are ways of behaving, interacting, valuing, thinking, believing, speaking, and often reading and writing that are accepted as instantiations of particular roles by specific *groups of people*, whether families of a certain sort, lawyers of a certain sort, churches of a certain sort, and so on through a very long list. They are always and everywhere social. Language, as well as literacy, is always and everywhere integrated with and relative to *social practices* constituting particular discourses. (Gee, 1990, p. xviii)

For example, what a teacher does in the classroom, even his or her attitudes, are part of the discourse (Johannesson, 1998). He or she learns the discourse by becoming a member of the group (of teaching). Starting as beginners, teacher education students watch what's done, go along with the group as if they know what is going on even when they don't, and eventually learn to do it on their own, with something of their own style. However, on becoming experts, teachers can't often say what they do, how they do it, or why, even though they can show someone (Gee, 1990). Being part of the discourse is to have "tacit knowledge" of how to behave (Gee, 1990; House, 1998).

Each Discourse incorporates a usually taken-for-granted and "tacit" theory of what counts as a "normal" person and the "right" ways to think, feel and behave. These theories crucially involve viewpoints on the distribution of "social goods" like "status", "worth" and "material" goods in society (who should and should not have them). (Gee, 1990, p. xx)

But at the same time each of us is positioned by many discourses and each discourse "represents one of our ever multiple identities" (Gee, 1990, p. xviii). And these discourses need not, and often do not, represent consistent and compatible values.

There are conflicts among them and each of us lives and breathes these conflicts as we act out our various Discourses. For some, these conflicts are more dramatic than for others. The conflicts between the home-based Discourses of lower socioeconomic black children and the Discourses of school are many, deep and apparent.... The conflicts are real and cannot simply be wished away. They are the site of very real struggle and resistance. (Gee, 1990, p. xviii)

The most powerful discourses in society have firm institutional bases (Weedon, 1987), for example in education. Because of this, these institutional settings are themselves “sites of contest, and the dominant discourses governing the organization and practices of social institutions are under constant challenge” (Weedon, 1987, p. 109).

From this perspective then, “it is the practices themselves, focused in technologies and innumerable, separate localizations, which literally embody what the analyst is seeking to understand” (Dreyfus & Rabinow, 1982, p. 187). And in seeing history in this way, the whole of Western science appears as historically constructed discourses that multiply and are variously located in relation to the “ruling apparatus” (Middleton, 1998, p. 116). Rather than history, and the events that have happened within it, being explained through some systematic global theory which holds everything in place, for Foucault, history is more a process without a subject, “or rather, it may be described from the standpoint of any number of subjects” (Patton, 1979, cited in Middleton, 1998, p. 116) where, for the moment, the “subject” is power itself. Using this notion of discourse and the notions of power explained above, I have argued in this thesis that the period in which this investigation took place was one of a number in which there has been a rupture in the dominant discourses of assessment since the late 1880s, when schooling became compulsory in New Zealand.

## **The Structure of the Thesis**

In order to understand the historical and social significance of discursive themes that emerged in New Zealand schools in the 1990s, I needed to provide an account of teachers’ assessment practices in the past. This was not easy. Histories of New Zealand compulsory primary education seemed mainly to be a view from the top (D. McKenzie, 1984; Middleton, 1987; Openshaw, 1987) and appeared to have little to say about assessment other than what the official policies had been (for example, Butchers, 1932; Ewing, 1960, 1970).

Reading further into the history, however, I came across views from the classroom – teachers’, students’ and inspectors’ accounts of classroom assessment (for example, Ewing, 1960, 1970; H. Lee & Lee, 1998; Matthews, 1987; Middleton & May,

1997). They were few and far between and difficult to access but enlightening as they revealed the contested nature of assessment throughout the time in which education has been compulsory in New Zealand. I came to see that my perception that there were many debates about the nature and process of classroom assessment in the present had also been the case in the past; and, in addition, that breaks and shifts in the policy had occurred causing teachers to change their assessment practices, beliefs, or both. Thus, Chapter Two uses a historical framework to trace the discursive themes of assessment of student learning from the beginning of compulsory education in New Zealand. While much of the material consists of “state educational discourse” (that is, curriculum statements, syllabi, policy statements and documents and official reports), I have also used teacher, student and school inspector descriptions in order to investigate the assessment practices of the past. In particular, I argue that teachers’ assessment practices are affected directly and indirectly by changes in social, economic, political and educational conditions and the discourses that accompany them.

Chapter Three examines the structural changes to education in New Zealand throughout the 1980s and early 1990s, and their effects in terms of assessment policy and practice. Evidence is presented to support the notion that in the latter half of the twentieth century there was a discursive shift from education as a human right to education as a government investment, and that this shift refocused public attention on both national standards and individual achievement. This chapter demonstrates how the official language of education and the structure of the education system and the curriculum changed and necessitated new technologies for measuring and monitoring performance. However, these official discourses of education were not the only forces at work on teachers’ assessment practices in the 1990s.

Chapter Four reviews the contemporary educational assessment literature and highlights further assessment discourses within which primary teachers must operate. International research carried out since the educational restructuring of the 1980s is highlighted and, in particular, the tensions that existed due to the demand for accountability and the need for assessment to improve learning are investigated. Several major assessment research projects and their results, particularly but not exclusively in New Zealand, are reported.

In Chapter Five the research questions that drove this study are presented. This chapter describes how multiple methods were necessary to investigate teachers' assessment practices in the wider national context, and, more specifically, in two Waikato primary schools and three classrooms within these two schools. The participatory case study design that used recursive strategies to circulate the data and their analysis for continual reappraisal, and the central place of the teachers as participants in this work, is advocated.

Chapters Six, Seven and Eight tell the assessment stories I heard from the teachers who participated in this investigation. Seeing teachers situated at the intersection of multiple and competing discourses of assessment was supported by the data from the interviews reported in Chapter Six. From the stories the teachers told about assessment and through an analysis of school and official assessment policies, it appeared that the teachers in the two Waikato schools were impacted by a maelstrom of ideas, theories, facts, judgements and expectations through which they had to navigate. In the context of self-managing schools, where the conflicting discourses of compliance with national regulations and the need to market education locally were dominant factors, the teachers I interviewed appeared to be prioritising summative assessment strategies over formative ones.

Chapter Seven develops the argument that a large degree of resonance existed between the assessment practices of the teachers in the two Waikato case study schools and the 20 teachers interviewed in schools throughout the country. This chapter uses Foucault's notion of "grids of specification" in the analysis of teachers' coding systems used to record the extent to which each student related to the norm. Although all of these teachers used what is known as criterion-referenced systems of assessment, the specification of the curriculum as achievement objectives arranged in a hierarchy of levels, and school policies coupled with surveillance mechanisms, resulted in a trend towards producing learners described in terms of their proximity to the various achievement objectives the teachers selected. Three more or less distinct approaches to assessment and differentiation are proposed from the data gathered and analysed in this chapter.



In order to confirm the characteristics of the approaches proposed in the previous chapter, in Chapter Eight the focus narrows to the close observation of three primary teachers in their classrooms. From these observations it was possible to describe in more detail how personal professional knowledge accumulates and influences individual teachers in their practice. The evidence in this chapter demonstrates how surveillance singles out individuals, both teachers and students, regulates behaviour and enables comparisons to be made. By revealing disciplinary power at work in these classrooms, I was able to come to some conclusions about why many teachers in self-managing schools have diverted their attention away from formative practices of assessment that promote learning (Black & Wiliam, 1998b), and complied with expectations that have driven their practice in more summative directions. Several assessment researchers use the explanation that “high stakes” are a major factor driving teachers to use summative assessment techniques (for example, Crooks, 1988; Gipps, 1994a; Nisbet, 1993). While acknowledging that the stakes are raised in a competitive educational market place, this analysis identifies disciplinary power within the micropolitics of the school as the mechanism through which teachers’ practices become summative, often in contrast to their knowledge and beliefs about the importance of formative assessment.

Chapter Nine develops the argument that teachers’ practices are powerfully impacted by multiple and competing discourses of assessment and education. Using the notions from Foucault’s “toolkit” introduced in Chapter One and amplified throughout this thesis, I theorise the ways in which disciplinary power harnesses the collaborative nature of teaching in self-managing schools under the gaze of senior teachers, principals, boards of trustees and the Education Review Office. As a result, rather than the primary purpose of assessment being that of improving learning and the quality of learning programmes (Ministry of Education, 1993a), in reality contemporary assessment practice is shown to be an instrument of governmentality. Exhorting teachers to make their assessment practices more formative in order to improve standards, learning, and to “close the gaps” (Gerritsen, 2000) between low- and high-achieving students and schools, will not be enough to shift teachers’ predominantly summative practice. Thus, it seems clear that increasing accountability measures further will do little to improve

standards. Even teachers with the best knowledge and understanding possible cannot resist the performance imperative when the stakes are raised for purposes of comparison.

## **CHAPTER TWO**

# **ASSESSMENT DISCOURSES IN NEW ZEALAND EDUCATION**

Instead of thinking that history once had a definite origin so deeply buried in the past that we have lost touch with it, so that we must now remember it afresh as the “real” starting point and purpose of humanity’s “being in the world”, Foucault suggests that history is differentiated and fragmented into particular discourses, and that each fragment (each discourse) has a threshold, a process of birth and an equally complex process of disappearance which can be analysed and described. (McHoul & Grace, 1998, p. 51)

The disciplinary effect of the examination in the form of assessment “introduces individuality into the field of documentation [because schooling] has to define the aptitude of each individual, situate his level and his abilities, [and] indicate the possible use that might be made of them” (Foucault, 1977, p. 214). As introduced in Chapter One, it is this disciplinary technique, used as a dividing practice, which produces scholastic identity on an individual basis (Meadmore, 1993). Although teachers have always made assessment judgements about how well each of the children in their care learn, the predominant use of the term “assessment” appears to be a recent addition to educational discourses (Black, 1998). In the past, terms such as “standards”, “testing”, “measurement” and “evaluation” were more commonly used to describe how judgements about learning were made. In seeking to understand the web of assessment discourses in the 1990s, I examined the research literature, past and present official syllabi and curricula, and official handbooks and guidelines, as well as other documentary evidence about teachers’ practices, for traces of these assessment discourses. This examination consisted of the analysis of statements to identify distinct structures of thought; that is, to identify the “set of rules” by which it is possible to know something, to know about what can be said, stated and thought and to recognise the conditions that constrain or enable what can be said (McHoul & Grace, 1998). I draw here from Foucault’s toolkit (Middleton, 1999), utilising his genealogical

method to explore the discourses that appear to have constructed assessment (McHoul & Grace, 1998). I take the view that assessment discourses have been constantly emerging and changing in New Zealand, influenced strongly by the ebb and flow of learning, teaching and assessment discourses internationally. In doing this I provide a history of the present (Middleton, 1998) to support the argument that predominant discourses of assessment are produced in particular social contexts and change according to the structural arrangements and wider social pressures of the time.

Before I begin, however, it is necessary to explain that rather than present this history as “a triumphal procession towards liberty and enlightenment” (Middleton, 1998, p. 5) I have attempted to portray the development of assessment in New Zealand schools as a series of “ruptures or breaks in the systems of reasoning (knowledge) that generate the principles of action” (Popkewitz & Brennan, 1998, p. 14). This is, to a certain extent, because there is not one continuous thread running through the development of teachers’ assessment practices, but rather, to use Wittgenstein’s term, “a thread made up of many fibres”, the strength of which resides “not in the fact that some fibres run its entire length, but in the fact that many fibres overlap” (Popkewitz & Brennan, 1998, p. 14). It is also because at, or around, each of these breaks in the systems of reasoning, teachers needed to change their assessment practices or beliefs, or both. I have chosen to examine sections of the development of school and classroom assessment within historical periods marked by change. In some cases change was due to legislation, the 1877 Education Act for example, while in others there were changes in regulations (the abolition of the Proficiency examination) or recommendations for change contained in official reports such as the *Commission on Education* (1962) (the Currie Report) and *Administering for Excellence* (Department of Education, 1988) (the Picot Report). Making these divisions according to historical periods marked by change, however, is not simply a matter of deciding on the dates when changes happened. Realistically, the periods overlap and interact, and have therefore been presented in an overlapping manner.

## **Before Compulsory Schooling: 1840-1877**

Prior to primary schooling becoming compulsory in New Zealand in 1877, each of the ten provinces was responsible for running an education system for itself. Each province had the authority to establish district schools under the supervision of local committees, to make grants-in-aid from public funds to supplement the school fees, and, in some places, to help meet the cost of erecting buildings (Ewing, 1960). These schools were beset with numerous difficulties, among which Ewing (1960) notes the difficulty of getting enough qualified teachers, the lack of teacher training, deficiencies in equipment and buildings and low school attendance. He also notes that there was considerable unevenness in the quality and breadth of teaching.

### **Testing Knowledge and Taking Places**

Until a national system was established in New Zealand, inspectors in each province monitored the teaching, learning and assessment practices of schools in their region. Little information about classroom assessment can be found as meagre records were kept, but the available evidence suggests that teaching was mainly modelled on schooling in England, Scotland and Ireland (Ewing, 1960). Ewing's description of pre-1877 schools in New Zealand contains a few first-hand accounts of schooling at that time. One describes the practice of physically placing students in order of achievement within the classroom.

In the school routine the class under instruction stood in a single line on the floor for the purpose of taking places. At the start each pupil took the place he or she had held at the close of the previous session, the dux of the moment at the right-hand end of the line, the second top on his left, and the rest in due order down to the unpraised mortal at the bottom of the class. Chances to retrieve or improve one's position in the line were constantly occurring. In reading, for example, while each pupil in turn read a passage aloud, a keen look-out for errors was maintained by all. Those below the readers of the moment had the prior chance of correcting, and whoever first called out a valid correction moved up above the erring reader. In cases where the lower section failed, the pupils above the reader were appealed to, and the top place then awarded to the first to supply what was

wanting. Oral spelling, the only kind practised, furnished an exciting contest. Only one trial was permitted, and the pupil who tentatively gave out “neice” (sic) with watchful eyes on the teacher’s face might snap out the correct spelling in the same breath, but it was of no use; an easy chance fell to the next scholar. (Ewing, 1960, p. 26)

As this passage indicates, there was an emphasis on constantly testing knowledge. In this example retention was rewarded physically and publicly through standing in a comparative position. Although evidence is scarce, it would seem that there were no ongoing written assessment records beyond the examination register, as the order of places was constantly changing and the pupils themselves remembered their places from day to day. Inspectors checked such records as there were (Matthews, 1987).

In fact, inspectors were essential to the functioning of the school system from the earliest days in New Zealand (Ewing, 1960; H. Lee & Lee, 1998; Matthews, 1987). They were the first itinerant education officers (Ewing, 1960) and “were expected to ensure scholarly standards were attained and maintained and that spending on education was carefully supervised” (Matthews, 1987). In fact Matthews goes further, to say that the inspectors were relied upon to be the “public’s educational watchdogs” (p. 74).

Even before national standards were introduced, the inspectors “took a leading part in building up small education systems and in trying to maintain reasonable standards of work in the classrooms” (Ewing, 1960, p. 77). “Standards” in this sense meant the specification of what students should be able to do at each year level. In some provinces this was minimal. In Wellington from 1874 for the second standard, for example, the requirements were

1. Easy narrative (from, say, 3rd reader).  
Repetition of poetry.
2. Passage for dictation from reader.  
Writing in Copy Books.
3. Spelling words of dictation exercise.
4. Arithmetic: Multiplication of money, weights and measures. (p. 71)

In other provinces standards were set out for each level in far more detail. For example, in Canterbury, based on the pattern of standards in England, there were standards prescribed for reading, spelling, writing, arithmetic, geography, history and grammar (Ewing, 1960).

Although inspectors worked in most provinces to carry out the necessary examinations to ensure that the standards were adhered to (H. Lee & Lee, 1998), this was a difficult task due to the nature of the colony at that time (Ewing, 1960). Many children got no schooling at all. Those that did only attended school for a few short years, and attendance was often irregular. The quality of teaching varied substantially and even the provision for inspection varied from district to district (Ewing, 1960). W.C. Hodgson, an inspector in Nelson, stated that there were two distinct theories of inspection: the “detective theory”, in which the inspector’s business was pretty much confined to reporting on whatever failures to attain the standards they discovered; and the “co-operative theory”, in which inspectors and teachers worked together to attain best results (Minister of Education, 1878). Compulsory schooling, introduced through the 1877 Education Act, was seen as the vehicle to regulate schooling throughout the country as a whole (Ewing, 1960) and establish national standards by which schools, students and teachers could be judged. In short, in order to produce higher-quality educational outcomes, assessment would be used to examine national standards (the expected norms) through regulation and the surveillance of the inspectorate.

### The Case for Compulsory Schooling

In addition to improving standards, however, the arguments used for making schooling compulsory included the struggle in favour of scientific rationality over religion adopted by forward and progressive thinkers, and, in Britain at least, compulsory schooling was offered as a solution to two major problems of the time: crime and pauperism (Walkerdine, 1984). In Walkerdine’s opinion, the popular understanding that bad habits were the cause of crime and pauperism led to the possibility of believing that compulsory education would

solve the nation's ills. "That is, by the inculcation of good habits, notably reading, in order, especially, to read the Bible" (p. 166).

Likewise Simola (1997) related the emergence of compulsory schooling in Finland to its ability to offer pupils the possibility of salvation in the form of a certificate of citizenship. Durkheim (Simola et al., 1998) also supports the idea that compulsory schooling in society in the nineteenth Century came to be seen as "an initiation ceremony that makes of the initiate an entirely new man, a new man and a citizen" (Simola, 1997, p. 1). The examination of standards thus quickly became the way in which individuals demonstrated that they were ready for citizenship.

Traces of these discourses can also be found in the move to compulsory primary education in New Zealand in 1877. Shuker, for example, argues that "there were many who stressed the social control function of public education" and that compulsory schooling would eradicate "rowdyism and larrikinism" (1987, p. 128). The Canterbury inspectors, in their 1876 report, wrote that they looked forward to a time when the primary school system would give teachers definite approved work to do, and which tested the results of their labours against a universal standard (H. Lee & Lee, 1998). The universal standard was seen as assurance that pupils in New Zealand's schools were receiving an education that would qualify them to take their place in society (Lee, 1987; D. McKenzie, 1987). Thus by 1877, standards and examinations to determine them by, and inspectors to implement and oversee them, were becoming established as the dominant assessment discourse in New Zealand.

### **Standards, Proficiency and Examinations: 1877-1937**

A national system of compulsory education and educational standards was established in New Zealand from 1877 as a result of the Education Act. Although the desire for a common, accountable prescription of work in all of the then colony's primary schools was reinforced by public opinion that favoured equality



of educational opportunity, there was debate on this issue. Uniform standards were used as the political justification, however.

The advantage of this system, the Minister of Justice Charles Bowen informed the House (of Parliament) in 1877 was that it provided a uniform means by which to evaluate the work of teachers and pupils in every part of the colony. (G. Lee & Lee, 1992, p. 22)

The curriculum, outlined by the Act, was broad and included reading, writing, spelling arithmetic, grammar and composition, geography, history, elementary science and drawing, object lessons, vocal music, and sewing, needlework and domestic science (for girls). Based on the English Revised Code, pupils passed from one standard to the next as the result of an examination by an inspector. In reality, though, these standards examinations tended to work against the full implementation of the broad curriculum.

### Examining Standards

During this period of national standards, teachers mainly taught to and assessed from the set standards. These were set out for each of the six years of primary schooling – Standard I to Standard VI – as statements for each subject. For example, the New Zealand Standards listed in the New Zealand Gazette, in September 1878, for Standard I stated

Reading– Sentences composed of words of one syllable, and common words of two syllables, to be read intelligently.

Spelling– Easy words of one syllable.

Writing– The small letters and the ten figures, on slate, at dictation.

Arithmetic– Counting and oral addition by twos, threes, fours and fives, up to 100; numeration and notation to 999; addition sums of not more than three columns, multiplication of numbers not exceeding 999 by 2, 3, 4 and 5...

Object lessons, Singing, Disciplinary Exercises, etc.– as prescribed in Regulation 9. (Ewing, 1960, p. 127)

Inspectors received precise instructions to visit schools twice a year, once for a general inspection (which became known as the surprise visit) and once to

examine the children according to the standards (Ewing, 1960). The teachers received ten days' notice of the examination visit and had to list all the children, and their classes, ages and attendance since the last visit in the examination register. The exams comprised both oral and written tests. The inspector recorded the result for each student in the examination with a dot for a pass and a zero for failure. Although the register's use was presumably taken over by teachers after the abolition of each standard exam and then Proficiency, this examination register does not seem to have been replaced until 1955 (Ewing, 1970).

### The Inspectors as the Public's Educational Watchdogs

The role played by inspectors in this system appears to have been pivotal. Primary schools in the nineteenth century in New Zealand were often small, rural and geographically spread. Teachers were not necessarily trained and pupils, particularly in rural areas, faced difficulties in attending school (Butchers, 1932; Matthews, 1987). The inspector's role in promulgating the standards following the 1877 Act required that he or she know how all the students in the region were performing. In this way, the inspector was responsible for the state of education in a specific Education Board. Matthew's case study of one inspector at this time describes how his positions enabled him to have this oversight.

As both secretary (of the Education Board) and Inspector, Hill was in a unique position to ascertain the true state of education in the province. As secretary he was required to attend all meetings of the Board, take minutes of its proceedings,...conduct the correspondence,...take charge of all papers, and documents,...and perform all duties as may be required. His inspectorial duties were also considerable. He was to inspect all Public Schools according to the regulations. (Matthews, 1987, p. 149)

This inspection included holding standards examinations.

The Education Act, and the regulations for the examinations in the standards, require that discipline, organisation, and classification shall be satisfactory in every school. They require that certain results shall be shown amongst the

children...whether the teacher be trained, or untrained, certificated or uncertificated. (Matthews, 1987, p. 149)

Individual students received a certificate showing which standard they had passed. Pupils were required by the regulations to exhibit the last certificate they had acquired on transfer from one public school to another (Ewing, 1960).

To pass the standards in reading, for example, children were examined on one book which parents were required to purchase. As a consequence of this policy, teachers required children to learn their books by rote, and inspectors reported how many for that year could quote word for word the contents of their book even, at times, while it was closed (Price, 1987). Inspectors also reported other matters, such as teachers who employed simultaneous class reading and alphabetic methods of learning to read. Debate took place among teachers about the narrowing effect of the standards and inspectorial examinations on the curriculum. For instance, J.C. Adams argued against the objections raised by fellow teachers in the *New Zealand Schoolmaster* (the first national educational journal) in 1885 which were similar to those raised in the 1990s.

The objections that have often-times been urged against frequent tests such as I have spoken of are easily summed up in the one objection, "they take away so much time from teaching"! But I think that is sufficiently answered when we see that all through the test examination there is running the training necessary for after life. The children are being rendered brighter and readier by the smart searching questions of an oral examination and are being made more self-reliant, more accurate and more methodical by learning to answer the written questions. At the same time we must not forget that while we are questioning them, either orally or otherwise, we are discovering either errors in our own teaching or errors in the children's work, and are thus enabled to advance slowly but surely on a firm foundation, without which all teaching is not only useless but absolutely hurtful. (Adams, 1885, p. 124)

### Measuring the Efficiency of Schools

Although, initially, the standards examinations were intended solely as a check upon the teachers' estimates of their pupils' abilities, the fact that children

who failed the inspectors' examinations were compelled to remain in their current standard for another year established a way of measuring and classifying not only individuals, but schools (Butchers, 1932; H. & Lee, 1998). H. Lee and Lee report that as early as 1879 the annual standards examinations were seen as a way of assessing the efficiency of a school. They quote the then minister of education's statement to support this.

Other things being equal, the best school in a district was the school which passed a larger proportion of children than any other in the district; and at a lower age; and a district was making progress if year by year the proportion of passes increased and the average age of passing became lower. (H. Lee & Lee, 1998, p. 6)

These discourses of standards and efficiency impacted upon the education system as a whole for nearly 60 years. Although the intention of the 1877 Act had been to provide primary aged children with a broad curriculum, in practice the cult of efficiency disciplined schools and teachers to focus on getting the children to pass the standards (H. Lee & Lee, 1998). Examination day became the educational "day of judgement" for both pupils and teachers, and the "adaptation of classroom work to the requirements of the standards was the greatest internal change in the schools in the first quarter-century of national education" (Ewing, 1960, p. 103). In fact, nearly everything was subordinated to the desire of a high percentage of passes (Butchers, 1932). Quoting from inspectors' annual reports to the education boards that were presented to Parliament and printed, Ewing (1960) cites the inspector Henry Hill thus

It can scarcely be realised to what extent the memory, to the neglect of everything else, is being employed in this district in preparing the children for the standard examination... Teachers are forced to use what to them is the only mode of escape from danger in the annual examinations, and the children, unable to complain of the stones given to them instead of bread, become the victims of, what is to me, a cruel and unnatural system of teaching. (p. 98)

H. Lee and Lee (1998) point out that many of the nation's newspapers capitalised on this preoccupation with examination passes as a measure of school

efficiency. They note that the media helped to popularise the results of the annual examinations that were calculated as pass rates.

By the late 1880's, high examination pass rates had become synonymous with school "efficiency" and "effectiveness", and all of the nation's education boards now calculated and reported their standards pass rates annually in terms of a pass percentage, calculated on the basis of the number of pupils who passed a given standard, divided by the number who actually sat the examination. Thus teachers and headmasters were appraised by the simple expedient of a pupil performance rating profile wherein the "best" teachers (and schools) got most, if not all, of their pupils through the standards examinations. (H. Lee & Lee, 1998, p. 8)

So popular was the word "efficiency" with politicians and educators through the early 1900s that the phrases "school efficiency", "civic efficiency" and "national efficiency" appear as the cover logos on the New Zealand Educational Institute's (NZEI – the teachers' union) National Education magazine up and into the 1920s (National Education, 1919).

In 1894, some 17 years after the standards had been implemented, regulations were introduced to allow head teachers to determine passes for Standards I and II. Slowly, each of the standards exams became internally assessed, but the Standard VI exam (Proficiency, as it was known) was not finally abolished until 1937 (H. Lee & Lee, 1998). As a result, the desire to achieve a high percentage of passes, most notably in Proficiency, led to undesirable "backwash" effects: the narrowing of the curriculum; retardation of students (holding them back in Standard V until they were thought capable of passing Proficiency or left school); and the use of examination driving, cramming and drilling (Beeby, 1986; H. Lee & Lee, 1998). Although there were recommendations to abolish the Proficiency exam (for example, the 1928 Lawson Syllabus Revision Committee strongly recommended its abolition and the introduction of a Primary Leaving Certificate), the Department of Education continued to endorse it and strengthened the uniform nature of the test. This they did by providing uniform written tests that could be marked from a set marking guide, thus making them appear more objective and reliable (H. Lee & Lee, 1998).

From the description of the conformance effect the standards had on the curriculum between 1877 and 1937 it would appear that classroom assessment was very much focused on checking on individual progress towards meeting each of the standards. Even though head teachers were able to determine progress from standard to standard, examinations continued to be conducted by schools three times a year in much the same fashion as the Proficiency exam was conducted by the inspectors in Form 2 (Ewing, 1970). Although the Cohen Commission in 1912 recommended that these be reduced to twice per year, this was not acted upon. In fact, even infant classes (those below Standard I) were examined rigorously. Due to the tyranny of the examination system, the flexibility of programmes and the broad curriculum foreseen at the passing of the 1877 Education Act did not materialise (Ewing, 1970).

### Harbingers of Reform

Against the popular demand for measuring performance through the standards examinations there were, increasingly, arguments for alternative methods of assessment that would promote the requirements in the 1877 Act that all schools cover the broad syllabus. Despite the fact that the inspectors implemented and oversaw the examinations in each school, many were also vocal critics (Ewing, 1960; Matthews, 1987). Hill, for example, claimed that the national standards acted against much of the good work achieved in schools prior to 1878 (Ewing, 1960; Matthews, 1987), and Vereker-Bindon of Wanganui and Lee from Wellington were scathing in their criticism of the growth of cramming for examinations (G. Lee & Lee, 1992). A number of education boards were anxious about the practice of keeping children in after school to prepare for exams (G. Lee & Lee, 1992), and correspondence, recorded in the New Zealand Schoolmaster journal, demonstrates that the NZEI was active in pushing for the abolition of standards against the wishes of some teachers.

From your editorial I gather that the Institute (NZEI)...has decided to memorialise the Department (of Education) to abolish the third standard history from the syllabus. I can't help thinking that such a move would be in the wrong direction. To me and my third standard scholars at all times the pleasantist hours of

each week have been those allotted to history.... What besides history, and to a certain extent grammar, have we to offer of an intellectual and enlivening character in the third standard studies? (Carlos, 1885, p. 124)

From the turn of the century more progressive ideas about education were actively introduced, at least at the policy level (Price, 1987). The reformist Inspector General of Education, Hogben, appointed in 1899, in explaining the new 1904 syllabus to a conference of inspectors, referred to Froebel and others of the most enlightened of the world's educators (Middleton & May, 1997) in stating that children will learn best, not so much by reading about things in books as by doing. Statements of this kind reveal that from early in the twentieth Century, even though practice was slow to change, there were crosscurrents of older and newer ideas at work in the educational landscape. Hogben, though, was pragmatic about the difficulties of implementing what were seen as kindergarten methods into infant classrooms (Middleton & May, 1997). In 1908, in a report to Parliament, he made a plea that the spirit of the kindergarten system, where children find out as much as they can for themselves by doing and making, discovering by their own trials (Middleton & May, 1997) be taken on by teachers in the early years of primary school.

As well as the child-centred discourses promoted by Hogben and others, progressive discourses were advanced by "scientific empiricists who directed their energies towards the test construction movement and looked to Francis Galton as their founding saint" (D. McKenzie, 1988, p. 76). Although far from the laboratories of England, France and the United States, New Zealand was soon influenced by intelligence testing and mental measurement. In 1920, the principal of Wellington Teachers' College presented an address to the Manawatu branch of the NZEI explaining how this type of testing had been used in England and the United States with some measure of success. He stated

- that these tests provided universal standard apart from instruction, environment, race, etc. We could compare French, British or New Zealand children;
- the tests were equally fitted for infants, children, adolescents, or adults;

- they eliminated the danger of the teacher's estimate being too high with older and maturer children and too low with younger children;
- they formed the most convenient and reliable method of testing our own examinations;
- and, covering a much wider field, they would serve as a much truer guide for choice of vocation. (National Education, 1920, p. 403)

The first use of mass mental testing in New Zealand occurred in 1924, when the Terman Group Test was administered to all entrants into New Zealand secondary schools (Moss, 1989). This efficient instrument was chosen because it was especially designed for testing “educability”, and the public and teachers were left in no doubt that the test was being investigated as a possible instrument for selecting students for streamed courses (D. McKenzie, 1988).

Together, these two factions of the early progressive movement – the child-centred theorists and the scientific empiricists of the mental-testing brigade – were packaged as “The New Education” (D. McKenzie, 1988; Middleton & May, 1997). But the alliance of these two schools of thought was an uneasy one (D. McKenzie, 1988). As Walkerdine (1984) explains, progressivism was made possible by specific sciences but was the result of a precarious congruence in which regulation, classification and liberation coexisted as simultaneous promises. “This unlikely coexistence was allowable because the true nature of the child to be liberated was guaranteed by the techniques of classification themselves” (Walkerdine, 1984, p. 164).

### **The Natural and the Normal: 1920s-1960s**

The post-World War 1 period was internationally a time of “venturesome” theorising and “bold” experimentation in curriculum and classroom practice (Ewing, 1970). In the 1920s and 30s these progressive discourses gradually began to impact on some classroom practice and on assessment. The “natural”, child-centred discourses were based on the idea of learning through doing (activity-based learning) and the natural development of the child; and the “normal” testing



discourses derived from scientific measurement of intelligence and achievement. Until 1937, however, when the Proficiency examination was finally abolished, standards examinations and the cult of efficiency still predominated (G. Lee & Lee, 1992).

### The Natural

Middleton and May's (1997) interviews with teachers reveal that from the 1920s the kindergarten and junior school curriculum was sporadically influenced by child centred and activity-based discourses even though it was still carefully planned and regulated by the teachers.

They had blocks to play with and they had a doll's corner. There would be books; they had a table of jigsaws, they had painting easels and crayons, but each activity was only for a certain period.... They had to stay a certain amount of time and work at the activity and then move on to something else. (Middleton & May, 1997, p. 31)

In school much of the reading matter in the 1930s was the *School Journal*. One interviewee cited by Middleton and May recalled that as a teacher she spent a great deal of her time hearing children read individually. "It was all individualised teaching. You kept the page a child was up to in a book and a bright child would say, 'Please can I read another page?' " (Middleton & May, 1997, p. 38).

Progressive educational ideas from overseas, while not sweeping the country, did receive support and caused considerable debate during the 1920s and 30s (Middleton & May, 1997). From England, Percy Nunn's *Education, Its Data and First Principles* claimed that the primary aim of all educational effort should be to assist students to achieve the highest degree of individual development of which they were capable (Ewing, 1970). Also in England, at Summerhill, A.S. Neill was challenging the traditional type of classroom education, and in the United States John Dewey was publishing *Schools of Tomorrow* and *Democracy and Education* (Ewing, 1970). Walkerdine describes a scene depicted in a textbook produced for nursery teachers in Britain in 1939 that epitomises these discourses.

The teacher-as-scientist is shown observing, monitoring, recording, classifying: the child-centred pedagogy legitimated. The teacher is depicted with the notebook, her gaze taking in the children's play – the powerful fiction of the continuous gaze of (dispassionate) observation: the teacher as all-knowing, all-seeing, knowing her children as individuals. (1984, p. 187)

In short, the crux of these ideas was that education was to do with the whole child; that the child's individual personality was of primary importance; that the child's needs and interests were more important than predetermined subject matter; and that individual motivation for learning rather than external pressure should be the basis of schooling (Middleton & May, 1997). But they were in sharp contrast with the prevailing standards and efficiency discourses still driving the mainstream system. Rather than soften the Proficiency exam requirements, the fact that passing it led to a free place in secondary education meant that more students than ever strove to pass it, and "the public criterion of a 'good' teacher and school was a high percentage pass rate" (G. Lee & Lee, 1992, p. 27).

This was officially recognised in 1919 when the numerical teacher grading system came into use with inspectors attaching considerable importance to Proficiency examination results when assessing the professional competency of teachers. Ministers of Education had also repeatedly echoed this sentiment. (G. Lee & Lee, 1992, p. 27)

Throughout the 1920s, as pupil numbers entering for Proficiency increased and economic depression deepened, the government "confirmed its policy that only the 'right' kinds of primary school pupils with the 'right' kinds of examination qualifications should be permitted to embark upon an academic secondary education" (G. Lee & Lee, 1992, p. 27). In this climate of scarcity the then minister of education did away with schools accrediting Proficiency in 1931 in a bid to tighten the "soft scholarship in the primary schools and a lax secondary school entrance policy" (p. 27).

Businessmen, conservative politicians, and the National Expenditure Commission (1932) were all united in their call for Proficiency to be a "stiff weeding out" examination for junior free places. (G. Lee & Lee, 1992, p. 28)

The examination requirements standardised the curriculum and the pace of classroom teaching in most schools (G. Lee & Lee, 1992; Middleton & May, 1997). Despite pockets of progressivist thinking and the criticism of many inspectors (Ewing, 1970), most teachers used cramming and rote methods to ensure exam success. As John Allan, a Teachers' College principal in the 1980s, explained at that time

You learned everything by rote. There was very little emphasis on comprehension. As long as you could reproduce it, that was what seemed to matter. The teachers were teaching to the Proficiency requirements, and so they had that in mind all the time. Their efficiency was equated with their success in getting people through the examinations. (Middleton & May, 1997, p. 43)

Progress at this time was communicated between schools by means of a transfer certificate stating the standard so far achieved by a child. It was replaced in 1930 by a progress card, kept for each child, which contained some cumulative scholastic and attendance data (Ewing, 1970).

Set against these conservative discourses, the abolition of Proficiency in 1937 following the election of the first Labour government in 1936 came as something of a shock to many teachers.

The abolition of the Proficiency examination, that rocked and shocked everybody because that was the standard by which they judged the teacher – how many Proficiencies did you get? How many Competencies (partial passes) did you get? It shocked some of the older ones. They nearly died, because when it was thrown out, they had nothing. (Middleton & May, 1997, p. 54)

But its demise opened the way for more progressive discourses to prevail. One major vehicle used to spread these more liberal views was the New Educational Fellowship (NEF) Conference, "the avant-garde of child centred education" (Ewing, 1970, p. 153), held throughout New Zealand in 1937. Organised by C.E. Beeby, Director of Education from 1940 to 1960, to address the fact that New Zealand's isolation was a handicap to public support for educational change, a number of eminent overseas educationalists, including Susan Issacs, stopped over

in New Zealand on their way to Australia to speak about their ideas at public meetings (The Commission on Education, 1962).

Primary schools were closed during the conference and nearly six thousand teachers enrolled. Never before in the history of the country had audiences of such numbers assembled to listen to discussions on educational topics. Reform was in the air. Parents and the general public caught the enthusiasm, the meetings overflowed, and “education” dominated newspaper headlines. (Ewing, 1970, pp. 193-4)

### The Normal

Concurrent with the introduction of “natural” discourses of child development and learning there were attempts to introduce the use of scientific measurement to assessment in New Zealand. School surveys were trialled by inspectors in an attempt to establish norms against which each teacher might compare the work and “educable capacity of each of his pupils” (N.R. McKenzie, 1930, p. 1). For example, in Taranaki, school surveys of this nature were carried out through the Education Office and published jointly by the North and South Taranaki Branches of the New Zealand Educational Institute in 1927 and 1929. The preface of the report states

This little book is published with the object of making a permanent record of the results of pioneering attempts to apply modern scientific methods to the measurement of the products of school education. ...The whole attitude of the teachers affords evidence of a very progressive outlook and of a keen desire to justify the claim that education is a science. The effect of this attitude is noticeable in many schools. It augurs well for the future of education in the Province. (N.R. McKenzie, 1930, p. ii)

The science of psychometrics (the measurement of mental states or processes) developed from the work on intelligence and, more specifically, intelligence testing (Gipps, 1994a; Gipps, 1994b; Glaser & Silver, 1994). Francis Galton was the acknowledged “father” of mental testing in Europe (Olssen, 1988). Galton’s views about human nature included believing intelligence to be a distinct biological force within an individual, that it was inherited and that it underpinned differential achievement within the social structure (Olssen, 1988).

Alfred Binet, a French psychologist, also made major contributions to the development of intelligence testing early in the twentieth century (Gardner, 1993; Glaser & Silver, 1994; Olssen, 1988; Perkins, 1995). Although, like Galton, Binet was not able to define intelligence satisfactorily, he saw it as a distinctive phenomenon and remained convinced of the need for an instrument to measure it. Rather than one ability, Binet thought of intelligence as “a salad of contributing abilities” (Perkins, 1995, p. 42). He saw assigning a number to intelligence as practical even though his view of intelligence involved a great deal more than the number indicated. He also contrived a way to measure a particular child’s performance against those of other children. Relating human performance to the well-known bell-shaped curve in the physical sciences, psychologists developed the curve on which, by convention, I. Q. scores came to be represented (Perkins, 1995). Thus it was assumed that mental qualities were distributed normally, as in the physical sciences. Rather than absolute performance against prescribed standards, norm-referenced tests were designed to produce “familiar proportions of high, medium and low scorers” (Gipps, 1994a, p. 5).

Since intelligence was seen as observable, innate and fixed in the same way as other inherited characteristics, it could, it seemed, be measured like other attributes. On the basis of the results of such testing, individuals were assigned to schools, streams, groups or tracks that were seen as appropriate to their intelligence (or ability, as it became to be seen) (Gipps, 1994a; Glaser & Silver, 1994). Despite some objections to the widespread use of tests to determine educational opportunity, testing for selection and placement became an institutionalised practice in American schools and society in the early twentieth century.

In New Zealand, psychometric tests were used in the name of social efficiency “to abolish the economic loss and social discontent from the hazardous attempt to fit square pegs in round holes” (Faulds, 1988, p. 93). J.F. Wells was the principal of Kowhai Junior High School, which, from the 1920s, used attainment tests to stream students into academic or technical courses (Faulds, 1988). Opinion, however, was divided about the merits of using these tests for streaming.

In particular, members of the Labour Party denounced their use. Peter Fraser, education spokesperson, argued vehemently that “if the tests were used to determine vocation, the children would be subjected to one of the greatest frauds and injustices ever perpetrated upon them” (Faulds, 1988, p. 92). The election of the Labour Party to government in 1935 probably mitigated against the widespread reliance on standardised tests such as those that came to dominate American education.

### Post-Proficiency Progressivism

Due to the abolition of Proficiency in 1937, primary schools were freed from measurement by percentage pass rates against the set standards as outlined above. Rather than standards exams dominating the educational landscape, there was space for the progressive discourses to find a firmer footing. As Beeby (1986) recorded, these progressive discourses were international and, as he put it,

education should at every stage be concerned with the child as a complete human being...and should not only be tolerant towards individual differences, but should adapt its methods so as to utilize these differences in the interests of both the individual and the common good (p. xxii).

Although the name “standard” was retained for the classes between Standards I and VI, these progressive discourses underpinned the idea that children should be promoted from class to class by age and stage rather than by exam results (The Commission on Education, 1962; Walkerdine, 1984). By observation and monitoring, teachers were trained to recognise the natural normalised stages of development worked out through empirical studies such as Piaget’s.

The new notion of an individualised pedagogy depended absolutely on the possibility of the observation and classification of normal development and the idea of spontaneous learning. It was the science of developmental psychology which provided the tools and in which the work of Piaget is particularly implicated.... Recognising such a movement is absolutely crucial to understanding how the present pedagogic common sense “facts” themselves have become concepts, structures – stripped of their content and located in individuals. It is the

work of developmental psychology which has made that move possible by providing scientific legitimation of a process of knowledge as development. (Walkerdine, 1984, p. 178)

*Education* (a magazine for teachers, published by the School Publications Branch of the Education Department) communicated to New Zealand teachers the scientific discourses of the normal and the natural through the publication of articles such as one by Cyril Burt on the “Subnormal Child” (1948).

By a subnormal child I mean one who deviates so far below the mental level of the average child of his age as to require special attention and provision.... In order to discover the dull or backward, we may rely principally on standardised tests. (Burt, 1948, p. 39)

Other issues of *Education* included photos of free activities, project work and intermediate pupils taking part in “pupil activity”. In an article on “What makes a good school?” H.C.D. Somerset (a Lecturer in Education at Victoria University, Wellington) explained that the best way for teachers to create a good school was by way of child study, and the Chief Inspector of Primary Schools, D.G. Ball, explained in yet another issue what “curriculum integration” meant.

Integration then is the product of three main forces – the child, the teacher and the subject-matter, and the first two are the more important. Once the emphasis is shifted from subject-matter to personality, then the integrating bonds elude the pen of the curriculum-maker and cannot be set down in any printed syllabus of instruction. They are the intangibles of education.... It is only the teachers who can integrate the curriculum in the truest sense. (Ball, 1948, p. 115)

Teachers interviewed by Middleton and May (1997), looking back on their training and work as teacher educators, were quite specific about these influences in the 1940s and 50s.

We did quite a slice on human development, which I thought was a very important part of it. If you’re going to teach, you’ve got to have some appreciation of the nature of the beast you’re dealing with at various ages.... So we did a substantial chunk of human development. A lot of Piaget. Bowlby was one of the ones I used as a source. Susan Isaacs...I did quite a bit on motivation. (Middleton & May, 1997, p. 124)

In those days you didn't play with the children – you were just observing and supervising them. (Middleton & May, 1997, p. 133)

As the Currie Commission pointed out, although the system of age promotion allowed far more curriculum breadth and flexibility than one in which progress was dictated by passing standards, it also had some drawbacks (The Commission on Education, 1962). With pupils within a class mainly promoted according to their age, children with a far wider range of attainment faced the teacher than was the case under the national standards system. This meant that teachers needed to develop ways of assessing and catering for children at different levels within each class. Many teachers were experienced in achieving this to a degree as they had successfully taught in sole-charge and small country schools. Grouping became the main mechanism for dealing with this new need for intra-class levels within primary classrooms.

For every group you had on the blackboard there was some written reading which was using the words you were teaching that day.... You had four or five groups. All this work was on the blackboard ready for the next day. (Middleton & May, 1997, p. 153)

In terms of assessment Beeby recognised that the direction a school system took was “profoundly influenced by the type of examination or other qualification that marks the end of primary and secondary schooling” (1992, p. 151). He cited the long dominance of the Proficiency exam over the type of primary education as evidence of this. His view was that the best primary teachers had detested the Proficiency exam but that less able ones had welcomed its “very limitations” (p. 152). He explained that the choices for the replacement of standards exams had been limited to inventing a new external exam or constructing a set of standardised tests of basic skills “that were just coming into fashion overseas” (p. 152). There had been some preliminary work done on designing and using standardised testing (for example Faulds, 1988; N.R. McKenzie, 1930) but the New Zealand Council for Educational Research declared it was not one of their priorities (Beeby, 1992).



By 1944 the Education Department had decided that instead of either of these options they would continue to rely on the judgement of inspectors to monitor standards using their control of the teachers' grading system and to set up "an imaginative programme of inservice training" (Beeby, 1992, p. 153) in order to guarantee the maintenance of standards through the new curriculum. This was a progressive solution indeed.

New ways of monitoring and recording were also needed to communicate progress and achievement from class to class and from school to school. This was achieved by means of a new confidential, cumulative record card, introduced in 1943, which accompanied a pupil throughout schooling, and showed personal characteristics and medical history as well as scholastic ability. In 1955 the Progress and Achievement Register, to be kept for each class, replaced the examination register, a lingering remnant of the standards exams (Ewing, 1970). In both of these types of record, achievement in the various subjects was assessed and recorded on a five-point scale, forming, in effect, a normal curve, thus perpetuating assumptions about the random distribution of abilities in the population.

The five categories are interpreted as meaning that out of a representative group of 100 pupils, five would receive the top rating of 1, twenty 2, fifty, or about half, 3, twenty 4 and five the bottom rating of 5. This assessment is to be made by the teacher in relation to all the children of the age concerned and not just in relation to the children in a particular class. (The Commission on Education, 1962, p. 262)

But, almost inevitably, there was not whole-hearted acceptance of the broadened curriculum, more child-centred approaches to teaching and learning or recording and reporting achievement normatively. Throughout this period allegations were made about declining standards in reading, writing and arithmetic (H. Lee & Lee, 1998) and these reached a peak in about 1958 (Ewing, 1970). In that year, the minister of education asked the senior inspectors to report to him on standards of achievement in English and arithmetic. In the same year standardised testing (possibly using the OTIS test designed in America) of 12-year-olds was also carried out by the New Zealand Council for Educational Research (NZCER)

and the results compared with those conducted in New Zealand in 1927, 1931, 1934, 1948, 1952 and also with overseas norms. Ewing (1970) suggests that this was, to that date, the most comprehensive survey that had ever been made of standards in New Zealand (p. 265). The results showed that standards in spelling and arithmetic had remained constant and were much the same as those in Australia and Britain. They reported the level of reading to be higher than ever before and possibly considerably higher than that of Australian and British children.

Despite these findings, the Department of Education was accused by those who, for one reason or another, opposed the changes to primary education of “sitting in judgement of its own theories” (Ewing, 1970, p. 266). The government of the day decided that the time had come for an independent evaluation of the education system and in February 1960 the Commission on Education was formally constituted (Ewing, 1970). The Currie Commission was asked both to evaluate the work of the existing education system and to recommend guidelines for its future development (H. Lee & Lee, 1998). It reported in 1962 in comprehensive fashion.

### Accountability Discourses

In terms of assessment, the Currie Report contained both an evaluation of classroom practices and recommendations for systematic standardised testing. While supporting the continuation of “normal testing of units of work and the regular surveys that teachers have always employed to evaluate learning” (The Commission on Education, 1962, p. 261) and cautioning against “over-testing”, the Commission saw that standardised tests should be administered as checks at the end of each stage of schooling. It recommended the establishment of checkpoints at the end of the third, sixth and eighth years of primary school (The Commission on Education, 1962). In this way, the Commissioners believed the checkpoints should “supplement the estimates of class teachers who were uniquely placed to take account of various factors affecting the ability and performance of pupils” (H. Lee & Lee, 1998, p. 23).

The Commission reported that it saw the “five point” system of recording achievement on record cards and progress and achievement registers “as well suited to record the progress of a pupil’s attainment within a school in five or more homogeneous groups” (The Commission on Education, 1962, p. 262) but it believed that it was “very difficult, if not impossible, for the majority of teachers ... to bring it into accurate relation to national norms”. The great variation of teachers’ estimates for students entering secondary school was presented as evidence for this view. Standardised testing, it recommended, would provide a guide to teachers for making these normative decisions. The report also recommended that the record card be amended to provide space to record the results of these tests, and remedial work be undertaken as a consequence of the results.

As well as monitoring individual achievement, the Commission recommended that the Department of Education should contract the NZCER to prepare and administer national standardised tests in the basic subjects at five-yearly intervals “in such a form as to allow valid comparisons of achievement to be made at particular points (Standards I, IV and VI) in the primary school curriculum” (The Commission on Education, 1962, p. 37). This was the “sole recommendation the Commission designed purposely to allay public disquiet”. In this way it believed that the primary school system would be accountable to the general public in terms of maintaining standards. In effect, standardised tests would increase the panoptic view allowing the public access to view schools’ results without relying on the inspectorate as the only window through which the view might be seen.

### **Assessment and Learning – the Rise of Criterion-referencing and Individualisation: 1960s-1989**

Following the release of the Currie Commission report there was public and professional discussion and a conference was held to address the Commission’s national checkpoint proposal (H. Lee & Lee, 1998). The

conference comprised teachers, school inspectors, and teachers' college and university lecturers. Although the importance of assessment information to confirm teachers' estimates on the five-point scale was acknowledged, the conference recommended that standardised tests should be made optional (rather than compulsory nationally) from Standard I upwards (H. Lee & Lee, 1998, p. 23). The NZCER produced the standardised tests – the Progressive Achievement Tests – that were distributed to all primary schools by 1969.

Teachers, however, continued to debate the accuracy of the five-point rating scale, and a number of committees of inquiry and working parties “gave close attention to ways in which to measure the achievement levels of New Zealand primary school children” (H. Lee & Lee, 1998, p. 24). Although three of these committees – Improving learning and Teaching (1974), the Royal Commission on Social Policy (1988) and the Ministerial Working Party on Assessment for Better Learning (1989-90) – all considered national monitoring of educational attainment, no moves were made by the Department of Education to implement such schemes. As H. Lee and Lee explain

To some extent, national monitoring of different areas of the New Zealand primary school curriculum was already occurring owing to our participation in some of the comparative surveys of educational achievement undertaken by the IEA (International Association for the Evaluation of Educational Achievement).... Further information...has also been available by way of the standardised Progressive Achievement Tests, developed (and periodically re-normed) by the NZCER. (1998, p. 24)

The progressive discourses that had influenced life in primary schools in the 1950s and 60s now began to filter upwards into secondary schools and even a few university classes. The young teachers and university students in the late 1960s and 70s were the “baby boomers” who had grown up in the progressive era, enjoyed the advantages of universal secondary education and were now involving themselves in sociological concerns such as the rights of ethnic minorities, protests against the Vietnam War, and a desire for radical change in cultural, class and gender issues (Middleton & May, 1997).

Education courses in teachers' colleges and universities included more sociological issues in their curricula and used such neo-progressive books as Postman and Weingartner's *Teaching as a Subversive Activity* (1969), Herbert Kohl's *The Open Classroom* (1969) and John Holt's *How Children Fail* (1974) as set texts. For example, at Wellington Teachers' College and Victoria University in 1969-71 student teachers were not only required to have a thorough knowledge of Piaget's developmental stages but were also taught about standardised testing and introduced to the relevant syllabi and the teachers' handbooks, which were based firmly on progressive discourses of learning and teaching (personal recollections). At Hamilton Teachers' College in the same years, Barbara Harold remembers

It was the days when things seemed liberal and all things seemed possible. When we left at the end of our three or four years there was the kind of feeling that the child was the epitome and everything you did was centred on the child. Developing their creativity, developing their independence and allowing freedom of expression. (Middleton & May, 1997, p. 273)

These child-centred discourses in the 1970s urged teachers toward more individualised approaches to teaching. While teaching using groups was the norm, texts and methods that encouraged teachers to focus on individual students, particularly in the basic school subjects, were increasingly popular. John Holdaway's *Independence in Reading: A handbook on individualized procedures* (1972) set out how to run an individualised reading programme and included extensive assessment activities that allowed the teacher to pinpoint, for the purposes of teaching, where each child might be in reading ability. Marie Clay's research into reading as the patterning of complex behaviour (1972) was published in 1972, and cognitive theories of learning such as those of Bruner in *The Process of Education* (1960), Piaget and Vygotsky began to influence notions of learning and teaching. Streaming was de-emphasised.

It was a time when mastery learning theories were being discussed and we thought we should apply them. And that certainly changed our role as teachers. (Begg, cited in Middleton & May, 1997, pp. 226-7)

Labour became the government for the third time from 1972 to 1975 and a liberal-progressive spirit in policy was activated through consultative conferences that continued into National's succeeding term (Middleton & May, 1997). The Educational Development Conference in its report *Improving Learning and Teaching* recommended that "broad national guidelines which would perform a function similar to that fulfilled by the present syllabuses" (Working Party on Improving Learning and Teaching, 1974, p. 132) be encouraged and supported. This document also supported evaluation as an essential aspect of curriculum.

Not only do administrators and principals need to be kept aware of the effectiveness of the curriculum and teachers informed of the progress they are making in achieving the specific aims of their course, but students want to know how they are doing so they can refine their goals, adapt their learning, and gain in self-knowledge. (1974, p. 132)

*Improving Learning and Teaching* included, for the first time in such a report, a complete chapter on assessment. Entitled "Assessment – Individual and National" this chapter defined the purposes of assessment, considered certificates and examinations, commented on teacher preparation for carrying out assessment, made recommendations about a national assessment programme, proposed the idea of evaluations of schools as institutions and made brief comment on school reports. In previous publications assessment had been commented upon within other sections, such as "Standards and Checkpoints" (The Commission on Education, 1962). This change in referring to assessment directly appears linked to other changes that were occurring during this period. As mentioned above, the broadening of the curriculum and the sharpened focus on individual students were accompanied by a shift in beliefs about the purposes of assessment. It was no longer perceived as mainly performing the accountability function in terms of maintaining standards or being used as a selection device; rather, the Education Development Conference's belief, strongly put, was that

the present ways in which assessment is used for these purposes is not in the best educational interests of the individual, or of society. We note with approval the increasing emphasis on the educative rather than the selective function of evaluation. (1974, p. 160)

### Improving Teaching and Learning or Maintaining Standards?

The recommendations from *Improving Learning and Teaching* included specific training for teachers in assessment principles and procedures and that this be made a compulsory aspect of pre-service teacher education. They also recommended more inservice training in assessment, that a system of national monitoring of standards be established, and that methods of reporting to parents be improved (Working Party on Improving Learning and Teaching, 1974). While not commenting specifically on the impact of the five-point assessment system still current in primary schools, the clear message in this report was that assessment should, first and foremost, help students to learn. This was a clear departure from previous similar literature. Although standards were still seen as important, the emphasis had shifted towards a further refinement in individualising instruction and connecting it closely with assessment.

The main function of assessment should be to indicate to the pupil the progress he is making and to indicate to the teacher the progress of his students and the effectiveness of his teaching.... teacher-judgement, whether or not part of formal assessment, should help students to learn, and help them to take responsibility for assessing their own learning thus enhancing their ability to learn outside formal educational institutions, and to form a realistic appreciation of their own capabilities. (Working Party on Improving Learning and Teaching, 1974, p. 159)

This trend towards educative and individualised assessment in New Zealand was, again, related to international discourses in education and assessment. The idea that nature would take care of a child's development was being replaced by a neo-progressivism (Middleton & May, 1997) that stressed the importance of the child's own activities, particularly activity of mind. Piaget, among others, argued at this time that the activities of a child were important as they fed back into the developmental sequence, making the child co-responsible for his or her own development. Likewise, the new discourse positioned the adult

as an active agent, able to guide and influence the actions of the developing child (Hultqvist, 1998; Walkerdine, 1992).

From the early 1950s assessment of educational achievement came to be seen as more useful than standards, intelligence or ability, specifically for the cycle of planning, teaching, learning and evaluation. Termed “educational measurement”, this approach aimed to devise tests which looked at a child as an individual rather than in relation to other children and to use measurement constructively to identify strengths and weaknesses individuals might have so as to aid their educational progress (Gipps, 1994a). Wood (1986) cites Glaser’s (1963) paper on criterion-referenced testing as a watershed in the development of this new form of assessment. Criterion-referenced assessment refers to the use of criteria (or standards) to which the performance of individuals is compared in order for judgements of achievement to be made. In Wood’s sense, then, educational measurement

- deals with the individual’s achievement relative to himself rather than to others;
- seeks to test for competence rather than for intelligence;
- takes place in relatively uncontrolled conditions and so does not produce “well-behaved” data;
- looks for “best” rather than “typical” performances;
- is most effective when rules and regulations characteristic of standardised testing are relaxed; and
- embodies a constructive outlook on assessment where the aim is to help rather than sentence the individual. (Wood, 1986, p. 194)

Wood’s definition is useful as it highlights changes in the direction of thinking about assessment from objectivity and measurement of representative performances for the purpose of comparisons between learners towards a more descriptive stance in which individual learners’ achievements are central. In looking for best rather than typical performances, optimal rather than standardised conditions for the student were considered essential and current cognitive learning theory was taken into account. For example, Vygotsky’s zone of proximal development became the point at which the teacher recognised the best performances possible and assisted the learner to produce them (Gipps, 1994a;



Torrance & Pryor, 1995; Wood, 1986). In this view of assessment, collaboration between the teacher and learner was seen as essential.

Teachers and researchers in New Zealand began to develop new techniques of assessment in order to target programmes more specifically to groups and individuals. In contrast to the comparatively rough guide provided by the external standards examinations and the psychometric and standardised tests of the late nineteenth and early twentieth centuries, these strategies enabled teachers to understand more intimately how each student was progressing. Teachers learned to use diagnostic strategies such as running records (Clay, 1979) and Informal Prose Inventories (Holdaway, 1972), to hold writing conferences with individual pupils (Graves, 1983) and to use pre- and post-tests in mathematics. Underpinning these practices were ideas about learning signalled by new words, such as “whole language”, “ethnomathematics” and “constructivism”.

Although these were new terms to most New Zealand educators the concepts and ideas behind them were familiar – especially to early childhood and infant teachers, and to the secondary school teachers of the 70s who had embraced “neo-progressive” ideals. Basically the focus was on “where the kids were coming from”. (Middleton & May, 1997, p. 322)

Gipps (1994a) and others (Department of Education, 1989c; Herman, 1992; Meadmore, 1995) have used the term “educational assessment” to describe this discursive shift towards using assessment primarily to assist learning.

The teacher has moved to centre stage as an actor in assessment rather than being a simple administrator of “better” tests devised elsewhere.... Because of these developments educational measurement has now become called more generally educational assessment; this is largely because “measurement” implies a precise quantification, which is not what the educational assessment paradigm is concerned with. (Gipps, 1994a, p. 10)

Concisely, educational assessment is based on the understanding that all children can learn, given the right input and motivation, and that attainment is not fixed but is amenable to change through teaching (Gipps, 1994a). New Zealand teachers and educators in this period generally concurred with the notion of educational assessment and accordingly found it almost hypocritical to have to

judge each student normatively using the five-point scale on school records, as they were still required to do.

However, debate continued. A more conservative National Party government took office in 1976 and economic recession fuelled public dissatisfaction about education (Middleton & May, 1997). The notion of standards was again raised for formal consideration in a report entitled *Educational Standards in State Schools* (Department of Education, 1978). Instead of supporting and extending the notion of individualised assessment, this report sustained the normative discourse and the assumptions about the psychological deriving from the physical world, stating

Students are not equally capable of learning, of reaching the same standards, or of coping with other tasks required of them at school. A few surpass their fellows at some tasks. A few lag behind. The majority of students cluster about a central point. This point is called the average.... Teachers work within a frame of reference when they are judging the standard of a student's work. They use grades or marks, in reports, progress cards and testimonials, to indicate the student's progress and nearly all of these rating scales are based on the "normal curve of distribution". This "normal curve" can be shown on a graph. It represents the way in which many abilities are normally distributed among any large group of people about the same age. (Department of Education, 1978, p. 15)

In addition, this report called for a return to assessment against measurable standards that would be capable of communicating to "the student, parents and prospective employers whether his (or her) level of achievement is at a good or acceptable standard" (Department of Education, 1978, p. 100). It recommended better training of teachers to diagnose student' learning difficulties and to recognise underachievers. But the report also acknowledged in its conclusion that some useful exploratory work on new ways of recording and reporting students' performances had been undertaken. It suggested that the Department of Education and the New Zealand Council for Educational Research intensify this work and that it be "closely related to in-service courses of teacher education" (Department of Education, 1978, p. 100).

Throughout the 1980s, teachers and other educators, led by the Department of Education, continued to trial different ways of recording and reporting student performances. As had become the style of curriculum review and development since 1943, the Department of Education continuously undertook what was termed “rolling revision” (Ewing, 1970). By 1978 this tradition was

one of close, continuing involvement by teachers as well as by subject advisers, officers of the Department of Education and others whose work is outside the classroom.... These extensive and lengthy consultations have an important consequence for the work of schools. Teachers have time to become familiar with what is new and to build it into their teaching programme, and...change in the schools themselves is an evolutionary process. (Department of Education, 1978, pp. 13-4)

In comparison with many overseas systems (and later events in New Zealand) this process of rolling revision meant that the revision of syllabi was a lengthy process. As a result of discussion and development, sometimes with virtually every school in the country (Department of Education, 1978), it could take a decade to finalise a new syllabus and, in the words of *Educational Standards in State Schools*, rolling revisions “do not result in sudden or unpremeditated curriculum changes; they do not impose curriculum change, but lead teachers towards change” (Department of Education, 1978, p. 14).

### Keeping School Records

Following the return of a Labour government in 1984, and probably as a result of rolling revision in the assessment and recording area and the influence of international educative assessment discourses, *The Curriculum Review* (Department of Education, 1986) recommended strongly that a new form of monitoring achievement, featuring descriptive recording of students’ progress, based on levelled criteria, be implemented and the five-point scale be abandoned (Department of Education, 1989c). In explaining the reasons for change the Department of Education described the progress and achievement register and progress cards as inadequate for “conveying information about a student’s

performance, and also as a guide for further teaching” (1989, p. 17). *Keeping School Records*, as the new assessment and evaluation guidelines were called, also argued that faulty assumptions underlay the use of the 1-5 ratings scale, that there was concern that the ranking system did not lend itself to the identification of students with learning difficulties, that research indicated that labelling students on a ranked scale could be counter-productive and that there was a “need to make more information available about a student’s educational level and progress at times of transition from one educational institution to another” (Department of Education, 1989c, p. 17).

The new system of keeping records was due to become compulsory for all schools in 1989 but was never officially gazetted. It was overtaken by the implementation of *Tomorrow’s Schools*. As will be explored in Chapter Three, depressed economic circumstances, combined with new economic and managerialist discourses, became extremely influential from the time of the fourth Labour government in New Zealand. Furthermore, from the 1980s, standards discourses were repopularised throughout the Western world (Gipps, 1994a). It is interesting to note, however, that despite never becoming official, *Keeping School Records* had been implemented in almost every New Zealand primary school by the early 1990s (Faire et al., 1994) and was still being used by 56 per cent of the schools surveyed in research into recording and reporting in 1997 (M. Hill, 1998a), demonstrating that teachers had taken on board the notions of criterion referencing and adjusted their practice in that direction even though the official policy had been to do otherwise.

## **Summary and Comment**

In this chapter, I have set out to demonstrate that the type of assessment constructed by teachers in each of the periods identified was directly affected by changes to the education system as a whole. The intention has been to show how multiple and competing assessment discourses have influenced public opinion, policy making and teachers’ practice simultaneously. I have argued that while at

any one time some discourses have predominated, they have always been in competition with others, and that the assessment practices teachers have used have often been different from the predominant discourses in policy and published literature. For example, although the inspectors' standards examinations had been abolished earlier in the twentieth century, schools were still keeping children back until they had demonstrated the skills and knowledge expected at each standard until after the Proficiency exam was done away with in 1937.

Although difficult to unravel and write about, these competing discourses, it is argued, make up the web of beliefs and practices that constitute the human environment. By using the theory of discourse, rather than considering history "as an uninterrupted continuity, the work of a sovereign human consciousness" (Smart, 1985, pp. 37-38)

Foucault decentres the sovereign subject and places the emphasis upon analysis of the rules of formation through which groups of statements achieve a unity as a science, a theory, or a text. In consequence the history of thought reveals, beneath continuities predicated upon the assumption of a sovereign subject, discontinuities, displacements, and transformations. (Smart, 1985, p. 37)

Because discourses comprise a set of widely held ideas that society relies on to make sense of the world (Benton, 1999), they are difficult to displace, and they continue while new ones are formed and in spite of policies and regulations designed to supersede them. Thus discourses are never static and rarely stable. "At any one time there may be multiple and competing discourses in a process of flux" (Benton, 1999, p. 2).

As this chapter has demonstrated, assessment discourses in New Zealand from prior to 1877 until the 1980s were constructed within the social, economic, political and educational conditions of their time, often in competition and always in flux. The standards and efficiency discourses of nineteenth century faded in the face of more scientific theories about natural development and psychometric measurement, but, as discussed, remnants of standards as well as notions of accountability persisted, particularly in times of economic depression, in competition with more progressive discourses.

These multiple and competing discourses exerted powerful disciplinary forces upon the teaching population. Hierarchical observation, normalising judgement and the examination (Foucault, 1977) were instrumental in the success of each of the assessment regimes that dominated teachers' practices in New Zealand primary schools described in this chapter. Even in the pre-compulsory era, teachers ranked students according to their responses to assessment questions, and, in turn, these teachers and their students were supervised by the inspectorate.

Following the 1877 Act, national norms, in the form of standards against which students were "known" were prescribed. The teachers and inspectors were as the watchers in the panopticon, where the director is the teacher.

In this central tower, the director may spy on all the employees that he has under his orders; he will be able to judge them continuously, after their behaviour, impose on them the methods he thinks best; and it will even be possible to observe the director himself. An inspector, arriving unexpectedly at the centre of the panopticon, will be able to judge at a glance, without anything being concealed from him, how the entire establishment is functioning. And, in any case, enclosed as he is in the middle of this architectural mechanism, is not the directors' own fate entirely bound up with it? (Foucault, 1977, p. 204)

The inspectors were required to report the examination results to Parliament, and through the publication of the efficiency (or otherwise) of schools in the newspapers, the fate of teachers was bound up with those of their students.

The gradual ascendance of the natural and the normal discourses of assessment was no less controlling than the earlier standards discourses, however. Rather than being inherently liberating, Walkerdine has argued that the forms of pedagogy necessary to maintain order and regulate populations "demand a self-regulating individual and a notion of freedom as freedom from overt control" (Walkerdine, 1992). She argued that to produce children who would not need to rebel – docile bodies who would become a self-disciplined workforce – the classroom became a laboratory where development could be watched, monitored and set along the right path. The overt examination produced by the standards assessment discourse was gradually replaced by covert, individualized monitoring, with continual judgements made against a less than specific norm. Attempts to

make the norm more explicit, through national standardised testing, were only partially successful in the progressive educational climate that existed up into the 1980s in New Zealand. Despite the educative assessment discourses of the 1970s and 80s (and the associated move to criterion-referenced assessment) that replaced the more progressive ones that preceded them, it was not until major structural adjustments to the education system as a whole at the end of the 1980s that expected national outcomes, and ways of measuring them, resurfaced with more effect.

In 1984 the incoming Labour government received from Treasury a document entitled *Economic Management: Brief to the incoming government* (New Zealand Treasury, 1984), which signalled Treasury thinking at the time “based on ‘New Right’ theory and strategies for improving the New Zealand economy” (Olssen & Morris Matthews, 1995, p. 1). In a subsequent brief (New Zealand Treasury, 1987) arguments were set out specifically for education that signalled belief in the need for a radical change to a view of education as a commercial investment. These economic and managerialist discourses of education, along with their effects on assessment, are explained and developed in Chapter Three. It is worth noting here, however, that there is a total absence of “culture” in any of these discourses. Despite the fact that the New Zealand education system included “Native” and, later, Māori schools, the human development and assessment discourses up until the 1980s were silent on Māori and the effects of culture.

# **CHAPTER THREE**

## **SELF-MANAGING SCHOOLS AND MANAGERIAL ASSESSMENT 1984-1995**

The rhetoric about “driving up standards” can be read as an attempt to realign education to the commodity values that increasingly define worthwhile or legitimate knowledge in post-industrial societies...the task of education is not the dissemination of a general model of life, not to transform students’ minds but to supply the system with the merchandise it needs in the form of information and skills. (Broadfoot & Pollard, 1998, p. 17)<sup>2</sup>

Thus, in the pursuit of greater accountability, government policies have produced systems of managerial surveillance and control that have fostered within educational institutions a culture in which trust is no longer taken to be the foundation of professional ethics. (Codd, 1999, p. 45)

In Chapter Two I argued that teachers’ assessment practices are affected directly and indirectly by changes in social, economic, political and educational conditions and the discourses that accompany them. This chapter examines the structural changes to education in New Zealand throughout the 1980s and early 1990s, and their effects in terms of assessment policy and practice. This chapter concludes by addressing the competing discourses that pervaded primary teachers’ “landscapes” (Clandinin & Connelly, 1995, 1996; Popkewitz & Brennan, 1998) during this period and may have influenced their practices at the time the fieldwork for this study was undertaken.

In the 1980s and 90s New Zealand primary teachers were subjected to new ruptures and breaks in their systems of reasoning and knowledge (Popkewitz & Brennan, 1998), changes that may have had particular effects on their assessment

---

<sup>2</sup> Quotation said by Broadfoot et al. to be from Gray (1990, p. 10); however, I have not been able to obtain this original paper.



action. Olssen and Morris Matthews (1995) described major shifts in both ideals and practices in New Zealand education that took place in a “fairly sudden manner and over a relatively short time span” (p. 3). These changes subjected the education system to the laws of the market, undermining the egalitarian ideals of the welfare state that had dominated educational provision in New Zealand for most of the century. The shift was mainly from a concern with equality of opportunity to a concern with technical notions of efficiency and consumer choice (Gordon, 1995; Olssen & Morris Matthews, 1995; 1997; Peters, Marshall, & Massey, 1994; Robertson, 1999; Strachan, 1997).

### **Economic Rationalism, Corporate Managerialism and the Marketisation of Education**

From the 1970s nation states internationally were engaged in a complex process of reshaping themselves due “to global and national economic changes, the demands of international competition, and pressures by social movements for responses to take account of both difference and local needs” (Robertson, 1999, p. 122). Due to these pressures, and responding to what at times has been referred to as the “dual crisis of accumulation and legitimation” (Dale, 1990; Grace, 1990; Lauder, 1990), powerful groups from governments and business interests (loosely referred to as the ‘New Right’) were able to redefine the terms of the debate in education, placing the blame for societal ills and economic crises on the education system (Apple, 1992; Dixon, 1999), in effect imposing a new set of discourses. Referring to these changes as “globalisation”, Robertson (1999), following Harvey (1989), took the view that they involved a dual process of

- widening and deepening capitalist relations on a world scale articulated around three interlinked and interacting features:
- A global market principle – where global market quality and price is imposed upon the domestic supply of goods and services. As a result, global competition becomes a powerful force in shaping domestic economic behaviour.
- The deregulation and liberalisation of financial markets, coupled with technological innovations, which increasingly

renders the world system one global system regardless of time or space restrictions.

- The principle of flexible accumulation means that the global market principle now affects all forms of production capacity, including areas that in the past – such as education - were not regarded as shaped by the commodity-exchange relation. Industries are now viewed globally, linked as a network of providers who compete with each other. (Robertson, 1999, p. 122)

Despite the fact that there was no groundswell of dissatisfaction with the education system (Kelsey, 1997) and that the 1982 OECD review of New Zealand education policy found that “to an extent greater than in some other OECD countries the parents, citizens, employees and workers of New Zealand appear to be reasonably well pleased with what is done for them in schools, colleges and universities” (OECD, 1982, cited in Kelsey, 1997, p. 219) the Treasury recommended change from 1984. Two documents recommended the changes and paved the way for major structural adjustments to New Zealand’s education system at the end of the 1980s. The first, *Economic Management: Brief to the incoming government* (New Zealand Treasury, 1984), laid the foundation for the first wave of “New Right” policies through a detailed analysis of the economy and a blueprint for its management (Kelsey, 1997). The second volume, *Government Management: Volume II* (New Zealand Treasury, 1987), provided a brief on education to the incoming Labour government that presented the existing education system as inefficient and inequitable (Codd, 1995; Kelsey, 1997; Lauder, Middleton, Boston, & Wylie, 1988; Olssen & Morris Matthews, 1995; 1997). The arguments that specifically related to education were that

- Education could be analysed in a similar way to any other service (p. 2);
- Education shares the main characteristics of other commodities traded in the market place, and that it could not be analysed successfully as a “public good” (p. 33);

- New Zealanders were too optimistic about the ability of education to contribute to economic growth and equality of opportunity (p. 8);
- Increased expenditure in education does not necessarily improve educational standards or equality of opportunity, or lead to improved economic performance (pp. 8, 18, 39, 130, 132, 141, 142);
- The nature of education should be determined by society's aims (p. 6); in particular, education should be more responsive to business interests and to the needs of the economy (p. 27);
- Not only has the education system not adjusted to changed circumstances, but it has performed badly despite increased expenditure on it (pp. 6, 16, 18, 140);
- The reason that the education system has performed badly is because teachers and the educational establishment have pursued their own self-interest rather than the interests of pupils and parents (pp. 37-38);
- The education system lacks a rigorous system of accountability, meaning there is a lack of national monitoring procedures or of any satisfactory ways of comparing the effectiveness of schools in order to account for the public resources employed (p. 108);
- Government expenditure on education has been too high and should be cut (pp. 37-38, 146);
- Government intervention and control has abetted the inflexible nature of the system and causes credential inflation, which promotes educational inequality. (pp. 37-39);
- Government intervention and control have interrupted the "natural" free-market contract between producer and

consumer with all that entails for efficient and flexible producer responses to consumer demand (p. 41). (Olssen & Morris Matthews, 1997, pp. 11-2)

As a result of these arguments, the Treasury recommended less state involvement and greater decentralisation of systems and decision making (Olssen & Morris Matthews, 1995; Strachan, 1997). These policies were consistent with globalisation (Robertson, 1999) and had the “primary objective of achieving a total rationalisation of the agencies of state, including educational institutions, to bring them into line with the policy prescriptions of free-market economics and corporate managerialism” (Codd, 1999, p. 45). In these times, it has been argued, economics became a “master discourse” or “what Lyotard (1984) calls a ‘metanarrative’, to which all developments in the policy sphere must adhere” (Fitzsimons, Peters, & Roberts, 1999, p. 37).

#### Economic Rationalism, Managerialism and the “Reform” of Primary Education

Codd (1999), too, argued that economic rationalism constitutes a form of discourse (p. 46), and, following Boston (1991), went on to explain that the theoretical underpinnings of public sector restructuring in New Zealand were derived from three main sources, namely: “Public Choice Theory”; “New Public Management” (managerialism); and the new economics of organisations, including “principal-agency theory” and “transaction cost analysis”. Gordon (1995) has also described the way in which economic rationalism and corporate managerialism have emerged from “Human Capital Theory” (p. 65). While not examining these here in great detail, it is useful to consider each briefly as they help in understanding how and why educational changes in the 1980s and early 1990s led to increased external accountability measures and a “culture of distrust” (Codd, 1999, p. 45).

The first, Public Choice Theory (PCT), arose in the United States of America during the 1960s and 70s and has as its key assumption that all people should be treated as “rational utility maximisers” (Codd, 1999, p. 46).

Furthermore, in PCT, all social phenomena are reducible to individual behaviour. Thus, there is no such thing as “the public good” or “society” as “moral preferences are merely the subjective choices of individuals” (p. 46). In this view, all people are assumed to be entirely self-interested, such that their behaviour is directed towards opportunism; thus, anyone employed by the state will seek to maximise their personal gains without the constraints and incentives of market mechanisms.

Thus, the state will grow ever larger and individual freedom will be reduced more and more. It is not difficult to see how the application of PCT produces policies aimed at restricting the pursuit of self-interest by professionals, such as teachers, whose work falls outside the domain of an economic market place. One of the mechanisms proposed by public choice theorists as a counter to the maximising behaviour of the public service providers is that of contestability. ... The assumption is that if a professional service is to be both efficient and responsive to consumer demand, there must be an ever-present threat of competition from other providers and concomitant opportunities for consumers to exercise market choice. (Codd, 1999, p. 46)

The second, New Public Management (NPM), assumes that there is a body of scientifically tested principles that can be applied to the management of any organisation to improve its efficiency and its effectiveness in achieving predefined outcomes (Codd, 1999). Such managerial theories are not new and can be traced to such systems of industrial management as “scientific management” or “Taylorism”, used to make railroad systems in the USA more efficient (Codd, 1999). In order to improve efficiency, however, there needs to be some measurable outcome so that gains can be identified. Thus, in relation to education, the notion of “quality” became critical. In order to demonstrate efficient performance, educational institutions needed to define quality in terms of key indicators or outcomes that could be measured and recorded. In the pursuit of quality and efficiency, schools “must engage in objective setting, planning, reviewing, internal monitoring and external reporting” (p. 47).

Combining ideas from these two theories (Public Choice Theory and New Public Management), the marketisation of education becomes possible.

Education has been transformed into a commodity: something to be “provided”, “traded” and “consumed” in a competitive marketplace. “Knowledge” has been replaced by “information” and learning “packages” are now exported and imported along with other goods and services. The marketisation of education preserves many of the key features of other positivist approaches; now, however, value is placed not just on that which can be measured and quantified but also on that which can be sold. (Fitzsimons et al., 1999, p. 40)

Other authors too have argued that changes in assessment internationally can be regarded as much as an attempt to commodify education as a way to measure standards (Kelsey, 1997; Peters, 1995). Elliott (1996, cited in Broadfoot & Pollard, 1998) contends that the new emphasis on assessment as a policy device is not even about raising standards, but rather plays a central role in “changing the rules which shape educational thought and practice. They are part of a language game which serves the interests of power and legitimates those interests in terms of the performativity criterion” (Elliott, 1996, p. 16, cited in Broadfoot & Pollard, 1998, p. 17).

The third theoretical strand within economic rationalism that has been argued as influential in bringing about the educational restructuring of the 1980s and 90s was that of the new economics of organisations (Boston, 1991; Codd, 1999; Gordon, 1995). From this strand comes the idea that all social, economic and political relationships are fundamentally contractual in kind. For example, the principal, as one party, contracts with the teacher, as the other party, for certain services and outcomes. This theory assumes that all parties to the contract have equal rights and capacities, but frequently this is not the case, raising the probability that the interests of one party will come into conflict with those of the other (Codd, 1999). Contractual arrangements require compliance with the terms agreed upon, or, as is the case with the curriculum in education, the regulations promulgated by the Ministry of Education. Accountability systems in the contractual environment are intended to secure the achievement of quantifiable results in a publicly demonstrable fashion (Blackmore, 1988).

Following the re-election of the Labour government in 1987 and, subsequently, the election of a National government in 1990, a raft of new reports, policy documents and Acts of Parliament confirmed economic rationalism and managerialism as underpinning public policy in general and education in particular (Olssen & Morris Matthews, 1995; Strachan, 1997). Other theses (for example, Dixon, 1999; Strachan, 1997) have analysed selections of these documents and argued that all of the identified influential documents had in common a commitment to achieving greater efficiency and greater equity through such policies as reduced state intervention, greater choice and competition, privatisation, user pays, contestability, decentralisation, devolution, managerialism, and contractual accountability and surveillance. These notions permeated all of the major documents relating to restructuring in the primary school and were essential elements in the New Right discourses promoted as necessary for the improvement of standards in all schools (Willis, 1994). But, as Kelsey (1997) points out, the three goals that underpinned the restructuring were contradictory. These were

a deregulated, artificially constructed education market, where education was a commodity subjected to the rigours of supply and demand; the development of a highly skilled and technologically literate population, which required a centralised, skills-based approach; and community participation, diversity and accountability, which demanded effective democratic input into decisions on policy, operations and resources. (Kelsey, 1997, p. 220)

These contradictory goals were applied to the education sector, including primary schools, through restructuring the entire system. In order to explain how this came about, I have summarised below those documents and reports that appear to have been most influential in promoting new assessment discourses and impacting primary teachers' assessment practices in the 1990s.

### *Administering for Excellence (the Picot Report) and Tomorrow's Schools*

Following the recommendation in *Government Management: Brief to the incoming government* (New Zealand Treasury, 1987), David Lange, the Labour government prime minister and education minister in 1987, set up a taskforce chaired by Brian Picot, a supermarket magnate, to make recommendations about how to reform the education system. The Picot Report, *Administering for Excellence* (Department of Education, 1988), advocated sweeping administrative reforms based on “choice, an assumption of individual competence, cultural sensitivity and good management practices” (p. 3).

In response, the government set out its policy to restructure the education system in *Tomorrow's Schools* (Lange, 1988). There were some minor changes to the recommendations suggested by the Picot Report, but in the main the thrust remained the same. The twin objectives of *Tomorrow's Schools* – school self-management and self-improvement (H. Lee & Lee, 1998, p. 26) – were set in legislation when the Education Act 1989 came into effect on 29 September. After the Act was implemented the Department of Education was abolished and the Ministry of Education set up. Each school became a basic unit of the administrative structure. Run by a community-elected board of trustees, each school became fully responsible for detailing its own charter which, although some clauses were mandatory, defined the purposes of the institution and the intended outcomes for students. Each school was now also responsible for staffing, allocating funds, and buildings and maintenance (Department of Education, 1988; Lange, 1988), although it was not accountable for how finances were spent. In terms of accountability, the board of trustees for each school was required to report regularly to its community on the objectives of the school's charter and on how well they were being achieved (Department of Education, 1988). An independent body – the Review and Audit Agency (later renamed the Education Review Office – ERO) – was established to ensure that institutions were accountable for the government funds they spent and for meeting the objectives set out in their charter.



### Tomorrow's Standards

Hard on the heels of *Tomorrow's Schools* came the publication in 1990 of *Tomorrow's Standards*, the Report of the Ministerial Working Party on Assessment for Better Learning, set up earlier under the Labour government. This document was the response of the Working Party to the terms of reference that said

The Working Party shall:

- (a) recommend to the government procedures for assessment which:
  - i) can monitor the effectiveness of the New Zealand school system on student learning;
  - ii) assess the effect of individual schools on students' learning achievements;
- (b) recommend ways of reporting on the above, taking into account different audience needs;
- (c) within the context of New Zealand's dual cultural heritage, advise the government on the possible effects of such assessment and reporting procedures for students, teachers, the curriculum, schools, employers and the wider community;
- (d) as a preliminary step to making the final recommendations, prepare a public discussion document:
  - i) outlining the main issues of assessment for better learning;
  - ii) describing a range of models and procedures for assessment. (Ministerial Working Party on Assessment for Better Learning, 1990, p. 5)

In accordance with these terms of reference the Working Party consulted widely on the issues listed above and produced a discussion document, *Assessment for Better Learning* (Department of Education, 1989a). *Tomorrow's Standards* was the most comprehensive report ever on assessment and standards in New Zealand schools and contained recommendations, later implemented, for national monitoring of standards, school self-review, external reviews of effectiveness, and a continuance of monitoring the progress and achievement of individual students. For the first time, New Zealand's dual cultural heritage was acknowledged in terms of assessment and reporting procedures. This may well have been due to

Treaty of Waitangi discourses, also prominent in the 1980s<sup>3</sup>. While theoretical ideas of economic rationalism, introduced above, influenced the recommendations – for example, the report recommended that schools be required to report on outcomes, performance indicators and added value – the Working Party set out nine guiding principles that it believed should guide assessment procedures. Interestingly, these principles reflected discourses other than those promoted by the New Right. They were

1. The interests of the students should be paramount. Assessment should be planned, implemented, and reported in ways which maximise benefits for students; while minimising any negative effects on them.
2. The primary purpose of assessment should be to provide information which can be used to identify strengths and guide improvement. In other words, it should suggest actions which may be taken to improve the educational development of students and the quality of education programmes.
3. Emphasis should be given to identifying and reporting educational progress and growth, rather than to comparisons of individuals or schools.
4. Every effort should be made to ensure that assessment procedures are fair to all.
5. Adequate involvement of all parties affected is essential to the credibility and impact of assessment processes.
6. Self-assessment is the appropriate starting point for assessment.
7. Careful consideration should be given to the motivational effects of assessment practices.
8. Appropriate assessment of the effectiveness of schooling requires sensitive attention to many factors.
9. In the assessment of intellectual outcomes, substantial attention should be devoted to more sophisticated skills such as the understanding of principles, applying knowledge and skills to new tasks, and investigating, analysing, and discussing complex issues and problems. (Ministerial Working Party on Assessment for Better Learning, 1990, p. 8)

---

<sup>3</sup> Although assessment and New Zealand's dual cultural heritage were linked in *Tomorrow's Standards*, it was not until late 1999 before the NEG and the NAG required boards of trustees to report on Māori achievement within each school.

The recommendations of the Working Party reflect the contradictory nature of these principles set in the context of the goals of marketisation and managerialism. For example, although the report recommended that the national education system be regularly monitored, the Working Party was of the opinion that national monitoring of achievement be carried out through light sampling rather than wholesale national testing. The recommendations pertaining to assessment in primary schools in *Tomorrow's Standards* are summarised below because they were implemented in primary schools and set the scene for the assessment practices of teachers described in this thesis.

### National Standards and Monitoring National Performance

*Tomorrow's Standards* supported the need for maintaining high standards of education in New Zealand. It reviewed different ways educational standards can be measured, and suggested that participation and retention rates, levels of achievement and the public perception of education standards should be nationally monitored. However, the Working Party cautioned that assessment alone could not lead to improved learning and higher standards.

It is the interweaving of curriculum, good teaching and assessment that ensures quality of learning. We must guard against an increasing preoccupation with assessment, as though it alone will set right perceived problems in the education system. (Ministerial Working Party on Assessment for Better Learning, 1990, p. 16)

The Working Party went on to remind its readers that “where the stakes...are high for students, teachers, or schools, there is evidence that schools teach to the tests, with the usual consequences of a significant narrowing of the school curriculum” (p. 16). The report cited American studies to support this point, but could well have pointed to the New Zealand experience between 1877 and 1937.

The need to monitor national performance was recommended, as it had been in 1962 (the Currie Report), in 1974 (*Improving Learning and Teaching*), and in 1988 by the Commission on Social Policy (H. Lee & Lee, 1998; Ministerial

Working Party on Assessment for Better Learning, 1990). The Working Party favoured a light sampling approach that randomly selected students at years 4 and 8 of schooling and cycled over a four-year period (Ministerial Working Party on Assessment for Better Learning, 1990) in preference to any form of externally-referenced national testing.

### Assessing the Effectiveness of Schools

*Tomorrow's Standards* devoted a complete section to assessing the effectiveness of individual schools. As described above, the reform of the administration of schools in 1989 had delegated the responsibility of accounting for the learning of students to the board of trustees and, through them, the principal and the staff. *Tomorrow's Standards* pointed out that “ the government, general public and the community of each school would require assurance that schools are performing well, and that any weaknesses are being identified” (Ministerial Working Party on Assessment for Better Learning, 1990, p. 24). Thus each school needed to be able to provide both accountability and performance information. As the report pointed out

The restructuring has considerable implications for assessment and the reporting of information. The previous structure used a complex system of regulations to control schools – controls designed to ensure that those involved with schools complied with standards perceived to be necessary for an effective education system. The new structure already provides for centrally controlled accountability and performance development of schools. The Working Party believes that this must be matched by locally controlled assessment. (p. 24)

To do this, the report suggested schools design their own “self-assessment” plan which would state school quality indicators with clear criteria defined to allow them to measure their effectiveness. Not only would schools be audited annually under this plan, but they would also review themselves at least once every three years to ascertain the effectiveness of their programmes and policies.

As *Tomorrow's Schools* explained, “Performance indicators is a term that has come from the world of accountancy.... They are, however, a useful

management tool for those held accountable for managing a school and the learning in that school” (Ministerial Working Party on Assessment for Better Learning, 1990, p. 26). Through this strategy of school self-assessment, as well as being monitored periodically by a team from the ERO, it was proposed that schools could monitor how well their students were performing, and thus how teaching, curriculum, resources and management contributed to students’ academic and social development and performance.

### Assessment of Students

In contrast to national and school-wide monitoring, this section of the report focused on the assessment of individual students. There were sub sections for the assessment of children in primary and secondary schools. The main thrust of the primary section of the report was to support the new Primary Progress Records (Department of Education, 1989c) and confirm that they should be implemented over the next three years. However, there was an emphasis on defined levels of achievement that foreshadowed the emergence of a new levels-based curriculum. The report stated

The new system emphasises assessment and evaluation based on levels of achievement rather than on comparisons between students. Descriptions of what each child can do form the basis of comment. A child’s progress and development is related to sets of learning goals which identify broad levels of achievement in knowledge, skills, and attitudes. The learning goals are designed to guide the development of the school’s own system of regular assessment and evaluation. Schools are responsible for developing and maintaining the system. (Ministerial Working Party on Assessment for Better Learning, 1990, p. 33)

The report gave examples of these broad goals and how teachers should record progress against them. It recommended that teachers collect assessment data and maintain folios of representative samples of work for each student (p. 36). In addition, there were recommendations for a formal check on the progress of every child in reading (in English and/or Maori) at age six, and reviews of progress in general as each child turned seven, nine, and twelve (p. 36). A child’s

family would be part of these reviews, and plans would be made about how learning should proceed for him or her on the results of the review.

The Working Party also recommended that assessment tasks that could be used by teachers to gain a deeper insight into a child's ability and attitudes should be developed and communicated to schools. Rather than only being in literacy and numeracy, which the report suggested could have a limiting effect on the curriculum, these tasks should cover the curriculum. It was felt they would "provide external reference points against which schools could compare children's progress, moderate teachers' assessments, and reshape school policy" (p.37). However, the Working Party made it clear that they "certainly (did) not envisage that they would take the form of a battery of nationally normed tests" (p. 37).

Contestability was also introduced with regard to the assessment of individual students. The Working Party stated that

Once national curriculum objectives are prepared and promulgated and it becomes known that assessment tasks would be welcomed, the Working Party has no doubt that educational agencies and publishing houses will become involved in their preparation. (Ministerial Working Party on Assessment for Better Learning, 1990, p. 38)

### Assessment and the Dual Cultural Heritage

The report made the point that little had so far been done to determine ways of assessment that would be appropriate to Māori. The report outlined the debate about Māori education that came through during its community consultation. It emphasised that "with regard to standards, we believe that attention must be directed to the retention of Māori students in schools, to the training of teachers and to the means by which students are assessed" (Ministerial Working Party on Assessment for Better Learning, 1990, p. 53). Defining the difficulties faced by Māori students, Wally Penetito was quoted as saying

If you are a Māori student, the more you achieve the more you are separated from your Māori peers. If you don't achieve, you get to keep your mates but then you can't get a job. You get deprived whichever way you turn. (p. 55)

Although stating that it was not the place of the Working Party to say what Māori people wanted, there were recommendations about a monitoring and accountability system, and that the minister of education formally set a mechanism in place to report to Te Iwi Māori on issues pertinent to assessment. However, there was still no acknowledgment that monocultural assessment practices may be problematic, and one set of achievement objectives was seen as appropriate for all New Zealand children.

#### Pre-service and Inservice Training in Assessment Techniques

Finally, the report made very specific recommendations about the need for teacher knowledge in the assessment area to be improved. The Working Party believed that due to the Primary Progress Records project there was now a set of well-defined methods for collecting assessment data and reporting progress and achievement. It set out very detailed topics for inclusion in both pre-service and inservice training in assessment.

#### The Achievement Initiative

In 1991, the Trade Development Board, supported by the New Zealand Treasury and the Reserve Bank, produced a report entitled *Upgrading New Zealand's Competitive Advantage*. Consistent with earlier Treasury documents, this report directly linked the lack of economic growth in New Zealand to the education system. The report argued that to recover economically New Zealanders in general and the education system in particular needed to change their attitudes towards competition and management, away from the progressive discourses of the 1970s and 80s towards knowledge and skills needed to upgrade the economy (Codd, McAlpine, & Poskitt, 1995; Crocombe, Enright, & Porter, 1991). This report has been recognised as the ideological basis for the Achievement Initiative launched by the National government in April 1991 as a key educational policy (Dixon, 1999; Peters, 1992).

The Achievement Initiative was the policy which informed most of the government-sponsored assessment and curriculum developments from 1991 (Willis, 1992). The policy identified three main elements

- The establishment of clear achievement standards for all levels of compulsory schooling, first in the basic subjects of English, mathematics, science and technology, and later other subjects;
- The development of national assessment procedures at key stages of schooling, by which the learning progress of all students can be monitored in those basic subjects;
- The allocation of resources to schools to meet particular learning needs. These needs may be those of underachieving students, or those of exceptional ability. The allocation of resources to schools will include provision for teacher development programmes. (Ministry of Education, 1991, p. 1)

These elements maintain the features of economic rationalism and the marketisation of education. The Initiative intimated that New Zealand standards were deemed to be below those of other nations and that implementing more extensive and systematic forms of assessment would produce higher standards (Codd et al., 1995). The strategies that were promoted to raise standards were the commodification of education through the definition of “clear achievement standards for all levels”, surveillance of all students (and, therefore, teachers) through monitoring, and government control of the allocation of resources for particular learning needs and teacher development. The policy contained in the Achievement Initiative paved the way for a new, levels-based curriculum.

### The New Zealand Curriculum

In April 1993, *The New Zealand Curriculum Framework* (NZCF) (Ministry of Education, 1993a) was published under a National government. It specified the seven essential learning areas and eight essential skills to be achieved by all students. Although New Zealand had had a national curriculum for primary schools since 1877, the NZCF, in combination with a series of national curriculum statements, became the official policy for teaching, learning and assessment in New Zealand schools covering all compulsory schooling from years 1 to 13. As



*Tomorrow's Standards* had foreshadowed, in order to use assessment to pinpoint individual progress and achievement in the way recommended, a new curriculum with national achievement objectives was necessary (Ministerial Working Party on Assessment for Better Learning, 1990, p. 38).

Just prior to the production of the NZCF, the first national curriculum document – in mathematics – was distributed to schools. This document indicated the structure of all the curriculum documents that followed it. Each curriculum document was divided up into strands (for example, in mathematics, there were six strands: number, measurement, geometry, statistics, algebra and mathematical processes), and within each strand, achievement objectives were described for each of eight levels of the curriculum. Each level related to approximately two years of schooling, except for the last three (levels 6, 7 and 8), which related to Years 11, 12 and 13 respectively.

Thus by 1993 New Zealand had a curriculum defined in levels with predetermined learning outcomes (achievement objectives) through which learning might be measured, recorded and reported. This clearly specified progression in each learning area is the hallmark of reformed national curricula (Elley, 1993b), and its installation as the method by which all legitimate learning would thenceforth be measured in primary schools marked the point at which economic rationalism and marketisation became a practical reality for teachers (Clark, 1998; Dixon, 1999). As Dixon puts it: “in this limited conception of teaching and learning all areas of learning are designed to produce the attainment of achievement objectives with effective teaching judged against the attainment of those outcomes” (1999, p. 25).

Apple (1996), however, suggests that this is not the only or major reason for implementing a national curriculum with predetermined levels of learning outcomes. He argued

Its major role is in providing the framework within which national testing can function. It enables the establishment of a procedure that can supposedly give consumers “quality tags” on schools so that “free-market forces” can operate to the fullest possible extent. (*Italics in the original*) (Apple, 1996, pp. 32-33)

While national testing was not introduced in New Zealand following the production of the new national curriculum as it was in Britain, the possibility remained under discussion throughout the duration of the research reported in this thesis. In order to demonstrate their compliance and effectiveness, however, schools were required to use the achievement objectives to monitor the progress of their students. National Education Guidelines were published in 1993 (Ministry of Education, 1993b) comprising the National Education Goals (NEGs) and the National Administration Guidelines (NAGs). These set out the responsibility of school boards of trustees, through principals and staff, to foster students' achievement by delivery of the New Zealand curriculum (Ministry of Education, 1994). Specifically, in terms of assessment, they required that

1. Boards of trustees must foster student achievement by providing a balanced curriculum in accordance with the national curriculum statements.

In order to provide a balanced programme, each board, through the principal and staff, will be required to:

- i) implement learning programmes based upon the underlying principles, stated essential learning areas and skills, and the national achievement objectives;
- ii) monitor student progress against the national achievement objectives;
- iii) analyse barriers to learning and achievement;
- iv) develop and implement strategies which address identified learning needs in order to overcome barriers to students' learning;
- v) assess student achievement, maintain individual records, and report on student progress. (Ministry of Education, 1993b, p. 4)

As well, each board of trustees was required to

- i) document how the National Education Guidelines are being implemented;
- ii) maintain an ongoing programme of self-review. (Ministry of Education, 1993b, p. 4)

These “guidelines” –commonly known as the NEGs and the NAGs – were, in practice, the regulations that schools came to be audited against by the ERO, and are therefore critical to understanding the context in which primary teachers

worked at the time of the study reported in this thesis. Rather than strategies that might lead to improved learning, such as continuous assessment during teaching as part of the teaching process, self-assessment by students and formative feedback to motivate learning, these requirements use words such as “achievement”, “monitoring progress”, “national achievement objectives” and “student records and reporting”. The emphasis is on the use of assessment for summative purposes, monitoring and accountability rather than on integrating assessment with learning or as a guide for teaching. The insinuation is that through accountability, learning and the quality of learning programmes will be improved. So, although the Ministry of Education introduced assessment into the re-formed curriculum primarily as a process for improving learning (Ministry of Education, 1993a, p. 24), government regulation and the high-stakes nature of the ERO reviews in self-managing schools enhanced the possibility that summative assessment might prevail over formative in the classroom. In order to assist schools to interpret these guidelines and implement assessment appropriately, the Ministry of Education published a handbook on assessment for teachers, *Assessment: Policy to Practice*, in 1994.

### The Education Review Office

The research reported in this thesis took place in New Zealand primary schools immediately following the changes outlined above. Schools had become self-managing and self-improving, accountable to their communities for their performance and the individual progress of each of their students. The inspectors, who in the past had provided support and advice to assist schools to implement new strategies, were no longer in existence. In their place, the Education Review Office had been established to audit and review the effectiveness of schools. Although initially based on a collegial, partnership model (Thrupp & Smith, 1999), the appointment of a new Chief Review Officer, Judith Aitken, changed this stance as the ERO “began to employ a variety of technocratic methods and its discourse came to reflect this” (Thrupp & Smith, 1999, p. 188). Consistent with public sector management strategies, it was discouraged from giving advice.

Rather, its role became one of surveillance, monitoring to see that schools were complying with the regulations and writing reports on each school's compliance and effectiveness (or lack of it). These reports often made media headlines, providing parents with information on which to make choices between schools.

Codd (1994) has argued that the ERO has been strongly influenced by the ideology of economic rationalism. Thrupp and Smith (1999) concur, arguing that the ERO thus promoted a technical decontextualised view of what makes an effective school through reporting performance in relation to outputs, contractual compliance and the attainment of specific outcomes. In fact, the fundamental purpose of the ERO and similar agencies has been described as "making the managers manage" (Robertson, Dale, Thrupp, Vaughan, & Jacka, 1997, p. 5). In their review of the ERO for the Post Primary Teachers' Association, Robertson et al. set out explicitly the contractual and compliance arrangements the ERO had with the government, and the way in which the ERO's reports, and the threat of these, came to operationalise the educational market (1997, pp. 5-7).

Despite criticism of the ERO (for example, MacDonald, 1997) and reviews of their performance (Austin, Edwards, & Parata-Blane, 1997; Robertson et al., 1997), the ERO remained a major feature of New Zealand's educational landscape (Thrupp & Smith, 1999). At the time the research reported in this thesis took place, the ERO reports were focused on the outcomes of education, providing schools with "assurance audits" and "effectiveness reviews" (Thrupp & Smith, 1999). Gordon (1995) made the point that by this time the charter each school had developed had virtually lost its function. The NEG's and the NAG's had replaced the charters, and Gordon described the ERO's functioning at this time as having

increasingly adopted a view of accountability in line with agency theory, and has recently extended its scope.... The focus is with compliance both to national legislation and regulations and with the school's own goals; keeping agents true to their stated aims. (Gordon, 1995, pp. 60-1)

## **Technocratic and Managerial Assessment**

A result of the changes to the education system examined above was the rapid escalation of technocratic assessment discourses (Blackmore, 1988; Comber, 1996; Dixon, 1999; Willis, 1994). Willis explained that a technicist model of assessment is based on the assumption

that learning is a linear process involving the acquisition of discrete predetermined pieces of information or clearly defined skills. The mastery of more complex outcomes rests on the successful learning of more basic skills or knowledge. Assessment procedures are based on positivistic assumptions whereby assessment is a neutral activity that is able to make precise and objective judgements about the intellectual capacity of individuals. (1994, p. 163)

Links with discourses of economic rationalism and corporate managerialism in education have been established with policies of educational assessment, both locally (Codd, 1999; Dixon, 1999; Willis, 1994) and globally (Lingard, Knight, & Porter, 1993; Meadmore, 1995). My interest, as a teacher educator and educational researcher, was in the ways in which technocratic and managerialist discourses impact on what teachers do, i.e. the effects of such discourses in actual classrooms. These effects – preferred pedagogies, enacted curriculum, assessment practices, recording and reporting strategies – are rarely studied and difficult to ascertain (Black & Wiliam, 1998a; Comber, no date). Yet they are critically important in understanding how policies, intended to raise standards and improve learning, work.

At the macro-level some studies have explored school and classroom reactions to the changes introduced in this chapter. Wylie (1994; 1998) and Harold et al. (1999) report that as a result of the changes in the last decade assessment has increased, teachers still use a wide range of assessment procedures and assessment results are more often aggregated in an attempt to demonstrate school performance than prior to the restructuring. At least two masters theses have also explored the responses of individual schools and teachers (Dixon, 1999; J. Hill, 1998a). But by 1996, when this study began, the effects of the reforms on

the micro-practices of teachers in classrooms had not been investigated in the New Zealand context.

It is important to investigate the effects of these discourses (and others) because it is precisely these practices that, Foucault contended, constitute an understanding of government (Meadmore, 1995). Government, in Foucault's sense

Describes... a certain way of striving to reach social and political ends by acting in a calculated manner upon the forces, activities and relations of the individuals that constitute a population. (Meadmore, 1995, pp. 9-10)

Applied to the practices of educational assessment, Meadmore (1995) contended that state policy making represents a calculated way of "knowing" a population in order to be able to manage it in terms of social and educational administration.

This does not necessarily constitute a repressive mode of control, nor, conversely, should it be seen as a source of liberation. Significantly, through the effects of what Foucault (1980) terms "power/knowledge", techniques of educational assessment provide knowledge which is important to government in accomplishing certain goals. It is precisely this productive nature of power in the Foucauldian sense that makes it important that assessment should be investigated for its social and political effects. (Meadmore, 1995, p. 10)

In order to bring about the commodification of education – the goal of governments in New Zealand in the 1980s and 90s – it was necessary to promote technicist assessment models (Willis, 1994) that incorporated technocratic, economic and bureaucratic forms of accountability (Codd, 1999; Dixon, 1999). Murphy and Torrance (1988) argue that technicist assessment is characterised by the administration of "formal tests and examinations" (p. 7), but the production of a levels-based curriculum containing outcomes against which students' progress must be regularly monitored, whether by formal tests, teacher observations or other means, also provided the conditions for technical modes of assessment (Dixon, 1999; Elley, 1993b).

Of prime importance in technocratic and managerial assessment discourses are the products or outcomes of education (Blackmore, 1988). As explained

earlier, these outcomes are publicly assessed and monitored. Introducing an outcomes-based curriculum against which teachers would monitor educational progress and “national assessment procedures at key stages of schooling, by which the learning progress of all students can be monitored” (Ministry of Education, 1991, p. 1) were essential to accomplishing New Right goals. The achievement of quantifiable results made it possible for schools to become publicly accountable to external audits by such agencies as the ERO. In this way accountability functions to discipline teaching through surveillance. In schools, as in other social institutions, the constant surveillance of educational outcomes and the means by which they are produced has regularised and routinized practices, “constituting a new discipline of norms and behaviour” (Meadmore, Limerick, Thomas, & Lucas, 1995, p. 403). Foucault’s metaphor of Bentham’s panopticon is a useful analytic for this means of control (Foucault, 1977). The panopticon, a prison design from the early nineteenth century, allowed complete surveillance.

At the periphery, an annular building; at the centre, a tower; this tower is pierced with wide windows that open onto the inner side of the ring; the peripheric building is divided into cells, each of which extends the whole width of the building; they have two windows, one on the inside, corresponding to the windows of the tower; the other, on the outside, allows the light to cross the cell from one end to the other. All that is needed, then, is to place a supervisor in a central tower and to shut up in each cell a madman, a patient, a condemned man, a worker or a schoolboy.... Hence the major effect of the panopticon: to induce in the inmate a state of conscious and permanent visibility that assures the automatic functioning of power. So to arrange things that the surveillance is permanent in its effects, even if it is discontinuous in its action; that the perfection of power should tend to render its actual exercise unnecessary...that the inmates should be caught up in a power situation of which they themselves are the bearers. (Foucault, 1977, p. 201)

Through constant surveillance, or the threat of it, discipline

breaks down complex multiplicities into simple units...carefully repartitioned in a basic cellular space: for each individual a place and for each emplacement an individual. It breaks down activities and actions into simple, momentary movements, thus allowing their control and ordering through routines and timetables. (Patton, 1979, cited in Middleton, 1998, p. 7)

Foucault believed that by the beginning of the nineteenth century, a “gaze” which incorporated the “mechanisms of hierarchy, surveillance, observation and recording was part of the disciplinary apparatus of government as it operated as a power/knowledge mechanism” (Meadmore et al., 1995, p. 402). Although not confined to technocratic models of assessment and accountability, mechanisms of hierarchy, surveillance, observation and recording have been instrumental in installing managerialist policies and practices in New Zealand schools since 1989. This form of surveillance, it is argued, produces certain forms of professional being; it lends emphasis to compliance and a certain lack of professional responsibility, rather than the reverse.

### **Summary and Comment**

This chapter examined the changes to education in New Zealand that occurred throughout the 1980s and early 1990s, and their effects in terms of assessment policy. The evidence presented in this chapter showed how international trends have affected assessment and accountability discourses. The last decade and a half has seen a discursive shift (in policy at least) from education as a human right to education as a government investment (Comber, no date). These most recent overseas influences have refocused public attention on both national standards and individual achievement. The changes in education policy, regulations and administration brought about by *Tomorrow's Schools* in 1989, and the introduction of the re-stated national curriculum in 1993, provided the context in which the research reported in this thesis took place. In this New Right context, the official language of education shifted to reflect an economic approach described by terms such as “inputs” and “outcomes”. New technologies for measuring and monitoring performance were promulgated and schools made more “accountable” for learning outcomes. These new policies regulated a clear focus on “knowing” the individual student in order to diagnose needs and promote learning.



However, these official discourses of education were not the only influences on teachers' assessment practices in the 1990s. As argued and demonstrated in Chapter Two, assessment discourses are often in competition and always in flux. Teachers take time to be convinced of the wisdom of new discourses and practices, and are also influenced by the products of research and their own tacit knowledge. At the time the research for this thesis began there was little advice available to teachers about how they might meet the new requirements imposed upon them, and from my interactions with teachers in 1995 there appeared to be an assessment frenzy taking place (Ramsay, Harold, Hill, Lang, & Yates, 1995). Rather than the supposed rational approach outlined in this chapter, at the chalkface assessment appeared to be a contested area, a cause of stress and confusion. Although it appeared that the curriculum had been spelt out more clearly than ever before and that the regulations had seemingly clarified and specified in simple language what schools were required to do, schools and teachers were seeking help more than ever before. Why had this area of teaching become so problematic for teachers? Chapter Four reviews the contemporary assessment research literature and raises further prominent assessment discourses in order to set the scene for the research itself.

## **CHAPTER FOUR**

# **COMPETING DISCOURSES IN EDUCATIONAL ASSESSMENT: THE RESEARCH LITERATURE**

In so far as teachers have to be involved in both formative and summative assessment, they are bound to experience some tension between their two roles. The former is private, focused on the learning needs of their pupils, which it must be their first concern to serve. The other has to reckon with the pressures and constraints that come from outside. National and regional high-stake systems create pressures for teachers to work within a framework which drives both their school policies and parental expectations, and the personal concern for the best interests of their pupils might have to take second place. The teacher has to hold the boundary between the different requirements of the two roles. (Black, 1998, p. 120)

Assessment devices are necessarily instruments of power. They are not, and cannot be, measures or indicators of some purely objective, independently-existing state of affairs. Rather they act to transform, mould and even to create, what they supposedly measure. (Broadfoot & Pollard, 1998, p. 18)

As I have argued in the previous chapters, classroom assessment is not a new phenomenon. There has always been a need to find out what children are learning in school. What has changed are the purposes of assessment, the ways in which it has been implemented, whether through examination, ranking, norm referencing, description or observation of classroom activity, and the uses to which the information acquired is put. As I have also shown, shifting societal discourses, debates, theories and policies have caused ruptures and breaks and influenced changes in educational assessment since compulsory schooling began in New Zealand. The structural changes made to the New Zealand education system from the early 1980s, and the new curriculum and accompanying regulations, it is argued, caused new ruptures in the discourses of assessment current in schools at the end of the 1980s and into the 1990s.

At the same time as the New Right discourses, examined in Chapter Three, were beginning to impact upon assessment practices in New Zealand, worldwide assessment discourses were also changing. The educational assessment research is reviewed in this chapter as a precursor to the examination of the ways multiple discourses of assessment came to compete for New Zealand primary teachers' attention. Research into the effects of these assessment discourses on teachers' practice, both internationally and in New Zealand, is then reviewed. This chapter concludes by raising, for investigation, the questions that guided the empirical research in this thesis.

### **Educational Assessment Discourses**

As indicated near the conclusion of Chapter Two, primary teachers' assessment practices in New Zealand, just prior to and during the early years of restructuring by Tomorrow's Schools', had been considerably influenced by international changes in the direction of thinking about assessment from objectivity and measurement to a more descriptive stance in which individual learners' achievements were seen as central (Gipps, 1994a; Torrance & Pryor, 1995; Wood, 1986). In her book arguing for a broader view of educational assessment in the place of a narrower testing interpretation of assessment, Gipps (1994a) maintained that assessment in education needed to be re-conceptualised in the 1990s. She argued that in contrast to psychometrics and an examination culture, a theory of educational assessment needed to include a much wider conception of assessment; one that included classroom (or teacher) assessment, standard tasks, coursework, records of achievement and practical assessment, as well as written examinations and standardised tests. Her definition included the use of formative, criterion-referenced and performance-based assessment, as well as summative and norm-referenced testing. Within an educational assessment paradigm, she suggested, assessment needed also to achieve a broad range of purposes, including supporting teaching and learning, providing information about pupils, teachers and schools, being used as an accountability device, and driving

curriculum and teaching (Gipps, 1994a). In effect, what Gipps describes in her book as educational assessment covers the whole range of assessment in schools today.

The widespread use of the term “educational assessment” is a very recent phenomenon, gaining momentum throughout the 1970s and achieving official status in several countries during the 1980s (Department of Education, 1988; Gipps, 1994a; Herman, 1992; Meadmore, 1995; Murphy & Torrance, 1988). But educational assessment is also a very loose term, interpreted to mean many things to many people. In contrast to Gipps’ (1994a) definition, Herman goes so far as to say “educational assessment is in a process of invention” (1992, p. 74). In her survey of the research literature about “good assessment”, Herman includes aspects such as open-ended questions, exhibits, demonstrations, hands-on experiments, computer simulations and portfolios as examples. Educational assessment is often referred to in the US literature as “authentic” assessment (Gifford & O'Connor, 1992; Shepard, 1991), “performance” assessment (Linn, 1994; Resnick & Resnick, 1992; Wiggins, 1989) and “educative” assessment (Wiggins, 1998). In all of these notions of assessment, however, the emphasis is on what happens in the classroom through the action of the teacher. As Gipps elaborates

the teacher has moved to centre stage as an actor in assessment rather than being a simple administrator of “better” tests devised elsewhere.... Because of these developments educational measurement has now become called more generally educational assessment; this is largely because “measurement” implies a precise quantification, which is not what the educational assessment paradigm is concerned with. (1994a, p. 10)

Crooks (1988), in a major review of international literature on the impact of classroom evaluation practices on students, legitimated the shift from assessment as standardised testing, and testing more generally, to a focus on classroom assessment activity. The research he reviewed indicated that

although standardised tests do have important and widespread effects under some circumstances...students spend vastly greater amounts of time engaged in classroom evaluation activities than in standardised testing. Furthermore, surveys of teachers and

students have consistently indicated that they believe the educational and psychological effects of classroom evaluation are generally substantially greater than the corresponding effects of standardised testing. (Crooks, 1988, p. 438)

Crooks' review went on to synthesise research that related to classroom evaluation (which he defined as evaluation based on activities that students undertake as an integral part of the educational programmes in which they are enrolled) as students. He concluded that, from the research to date, classroom evaluation appeared to receive less thought than most other aspects of education but affected students in many different and important ways. For example, it guided students' judgements of what is important to learn, affected motivation and self-perceptions of competence, structured their approaches to personal study, consolidated learning and affected the development of enduring learning strategies and skills. In short, Crooks stated that classroom evaluation "appears to be one of the most potent forces influencing education. Accordingly, it deserves very careful planning and considerable investment of time from educators" (Crooks, 1988, p. 467).

In their work on educational assessment and classroom evaluation both Gipps and Crooks heralded more than a technical shift from testing to, for example, observational activities (Crooks, 1988; Gipps, 1994a). Gipps stated "that our underlying conceptions of learning, of evaluation and of what counts as achievement are now radically different from those that underpin psychometrics" (1994a, p. 158). She continued that

educational assessment recognises that domains and constructs are multi-dimensional and complex, that assessing achievement is not an exact science; and that the interaction of pupil, task and context is sufficiently complex to make generalisation to other tasks and contexts dubious. (p. 159)

According to Hattie, in educational assessment clear standards are set for performance against which pupils are assessed, feedback to learners is emphasised, and students are encouraged to monitor and reflect upon their own performance (Hattie & Jaeger, 1998). In this paradigm, students are encouraged to think for

themselves about important and relevant subject matter rather than to tick alternatives or regurgitate knowledge. In this type of assessment, best performance is elicited in authentic contexts and in non-threatening conditions.

In educational assessment we move away from the notion of a score, a single statistic, and look at other forms of describing achievement including “thick” description of achievement and profiles of performance. (Gipps, 1994a, p. 160)

Therefore, in the educational assessment paradigm, teachers’ own assessments of students are a key component (Gipps, 1994a; Herman, 1992). In order to make judgements, assessment needs to be an interactive process between teachers and learners so that understanding can be gauged and learning can be scaffolded (Kamii, 1982; Sadler, 1989). In order for this to happen, teachers need to know and understand the subject matter well, they have to know how to get at the learners’ knowledge and understanding and they have to know how to elicit best performances. This requires that educational assessment be a “low-stakes” activity (Sebatane, 1998). The publication of test data of individuals, classes and schools raises the stakes and distorts the educational process (Black & Wiliam, 1998a; Gipps, 1994a). As Crooks (1988) noted, too much emphasis has been placed on grading and ranking of students. Assessment in a low-stakes programme, where individual feedback is constructive and focused on an individual’s performance related to his or her own prior performance, is more likely to maintain the engagement of students, retain them in the education system and lead to improved standards (Crooks, 1988; Gipps, 1994a).

This description of educational assessment does not deny that educational standards are important. They are, but the literature indicates that improved educational standards will come about not so much through national testing or accountability measures, but through changes in the teaching/learning interaction in the classroom (Black & Wiliam, 1998a; Crooks, 1988; Hattie & Jaeger, 1998; Herman, 1992; Wolf, LeMahieu, & Eresh, 1992).

Ultimately, the realisation of these national policies on standards and assessment will not be played out in Washington, but in the way the teacher answers a child’s question; the way a principal finds joint planning time for teachers; and the way a community

of parents and teachers joins forces to see that these standards become a tool for, rather than an obstacle to, equity. (Wolf et al., 1992, p. 9)

Another major review of the assessment literature (Black & Wiliam, 1998a) criticised the inputs and outputs model of raising standards that was promoted in several Western countries, including New Zealand, in the 1990s. Black and Wiliam (1998b) argued that this approach has tended to treat classrooms like a black box.

Certain inputs from the outside are fed in or make demands – pupils, teacher, other resources, management rules, and requirements, parental anxieties, tests with pressures to score highly, and so on. Some outputs follow, hopefully, pupils who are more knowledgeable and competent, better test results, teachers who are more or less satisfied, and more or less exhausted. But what is happening inside? How can anyone be sure that a certain set of new inputs will produce better outputs if we don't at least study what happens inside? (p. 1)

Educational assessment, a classroom-based activity with the teacher and student at its centre, has emerged as a crucial technology for improving learning and raising achievement (Black & Wiliam, 1998a). As a result of their review of the assessment research literature in the previous ten years, Black and Wiliam revealed that

There is a firm body of evidence that formative assessment is an essential feature of classroom work and that development of it can raise standards. We know of no other way of raising standards for which such a strong *prima facie* case can be made on the basis of evidence of such large learning gains. (Black & Wiliam, 1998b, p. 19)

Formative assessment was thus singled out in this review, and by other researchers (for example, Bell & Cowie, 1997; Harlen, 1998; Sadler, 1989; Torrance & Pryor, 1995), as the aspect of educational assessment and classroom evaluation most critical to improving learning. Formative assessment provides feedback to teachers and pupils about on-going progress in learning. It has a direct influence on the quality of pupils' learning experiences and, thus, on achievement.

In contrast, summative assessment is the means for communicating the nature and level of pupils' achievements at various points in their schooling. Within the summative category at least three purposes can be noted: assessment for reporting achievement and progress; assessment for certification, used to summarise achievement for the purposes of selection and qualification; and assessment that provides the information used in judging the effectiveness of teachers, educational institutions and of the system as a whole. This last type has been called "evaluative" assessment or assessment for "audit" and/or "accountability" (Codd, 1999; Dixon, 1999; Wiggins, 1998) and "quality control" (Harlen, Gipps, Broadfoot, & Nuttall, 1992). Presented in this way, as they are in *The New Zealand Curriculum Framework* (Ministry of Education, 1993a) and in the assessment handbook for primary teachers in New Zealand (Ministry of Education, 1994), formative and summative assessment appear to have quite distinct purposes. However, the difference may not be so clear cut (Harlen, 1998). While the headings "formative" and "summative" have been used below to discriminate between these assessment purposes, much of the research reported within these sections demonstrated how these purposes can and do overlap.

### Formative Assessment

Although slightly varied definitions of formative assessment are offered in the literature, there seems to be general agreement that assessment becomes formative when "the evidence is actually used to adapt the teaching work to meet the needs [of learners]" (Black & Wiliam, 1998b, p. 2). Formative assessment is directly related to cognitive processes (Dixon, 1999; Wiggins, 1998), involves teachers in recognising and responding to student learning in order to enhance it during learning (Bell & Cowie, 1997; Glover & Thomas, 1999; Hattie & Jaeger, 1998) and is concerned that feedback on learning to learners is critical (Crooks, 1988; Harlen, 1998; Sadler, 1989; Torrance & Pryor, 1995; Wiggins, 1998). Sadler (1989) went further, stating

for students to be able to improve, they must develop the capacity to monitor the quality of their own work during actual



production. This in turn requires that students possess an appreciation of what high quality work is, that they have the evaluative skill necessary for them to compare with some objectivity the quality of what they are producing in relation to the higher standard, and that they develop a store of tactics or moves which can be drawn upon to modify their own work. (p. 119)

Sadler's explanation of the implications of formative assessment highlights a conception of student development as multidimensional rather than sequential, where prerequisite learning cannot be conceptualised as neatly packaged units of skills or knowledge, as seems the case in some other conceptions of formative assessment (Torrance & Pryor, 1995). Underlying these differing conceptions of assessment are very different perspectives about learning. Torrance and Pryor (1995) explain one perspective as closely related to the mastery learning tradition – a competence model – where objectives are defined and taught to quite specifically. The other perspective is seen as deriving from the social constructivist perspective in cognitive psychology (Dixon-Krauss, 1996; Kamii, 1982; Torrance & Pryor, 1995), where the role of the teacher includes “assisting the student to engage with and comprehend new ideas and problems (Torrance & Pryor, 1995, p. 308).

### Convergent Assessment

Where competence conceptions of learning underpin assessment, there is an assumption that all learning objectives can be defined and specifically taught to. In this view, objectives or learning outcomes can be pre-specified and arranged hierarchically, and teaching and learning proceeds through them in logical order. New national curricula in New Zealand and elsewhere appear to be underpinned by such theories of learning (Dixon, 1999; Elley, 1993a; Elley, 1993b; Filer, 1995; Torrance & Pryor, 1995). Torrance and Pryor (1995) argued that formative assessment, when underpinned with this view of learning, resulted in convergent teacher assessment. In convergent teacher assessment the important thing is to

find out if the learner knows or can do a predetermined thing. It is characterised by

detailed planning, tick lists and can-do statements, an analysis of the interaction of the child and the curriculum from the point of view of the curriculum, and closed or pseudo-open questioning and tasks. (Torrance & Pryor, 1995, p. 315)

This view of learning as incremental and hierarchical relates closely to Bernstein's description of a collection-type curriculum, in which "the contents are clearly bounded and insulated from each other" (Bernstein, 1971, p. 49). In such curriculum, Bernstein argued,

social order arises out of the hierarchical nature of the authority relationships, out of the systematic ordering of the differentiated knowledge in time and space, out of an explicit, usually predictable, examining procedure. (Bernstein, 1971, p. 63)

Such a curriculum model emphasises performance that requires a specific output from the acquirer, a particular text that the learner is required to construct and the acquisition of the specialised skills necessary to the creation of the required output (Broadfoot & Pollard, 1998).

Through the operation of explicit assessment procedures, learners will be made aware of the learning outcomes that will be valued. Their performance will in turn be a means of locating them in terms of a hierarchical judgement. (Broadfoot & Pollard, 1998, p. 14)

Broadfoot and Pollard (1998) believed that these features were readily recognisable in the schools participating in their research project, Primary Assessment, Curriculum and Experience (PACE), in English primary schools. Also in England, Gipps, Brown, McCallum and McAlister (1995), in their research into the effects of national assessment of seven-year-olds on teachers' practice, identified a group of nearly 50 per cent of their sample of teachers as evidence gatherers. These teachers conscientiously tried to adapt their practice to the national curriculum, prioritised teaching over learning, knew in advance which

objectives they would be assessing, and often had prepared can-do lists and tick sheets. Assessments mainly took the form of work set for pupils, which was collected and stored, often in individual portfolios.

When these teachers came to a point when they needed to record their assessment, they assembled together all their evidence and tried to use it to make a judgement against the criteria. (Brown, no date, p. 8)

Research evidence from Canada (Bachor & Anderson, 1994), Australia (Comber, 1996) and New Zealand (Bell & Cowie, 1997; Calder, 1995; Dixon, 1999; J. Hill, 1998a; M. Hill, 1998a) has indicated that similar assessment practices have been observed. Gipps et al. (1995) argued that this evidence gathering was summative, not formative, and Harlen (1998) concurred that “even ‘ongoing’ or ‘continuous’ assessment does not necessarily mean that the information is used to help learning” (p. 2). Rather than being an integral part of the teaching and learning process, this type of assessment takes the form of a series of short summative tasks or tests. Other researchers describing similar assessment practices, have labelled such teacher behaviour as “weak formative” (Brown, 1996), “informal summative” (Harlen, 1998) and “technicist” (Willis, 1994). It could be argued that “planned formative assessment” (Bell & Cowie, 1997; Cowie, 1997), where teachers used various planned methods to check on progress and record the results for later use in planning future activities, was also a form of convergent assessment.

### Divergent Assessment

In contrast, formative assessment underpinned by the social constructivist perspective in cognitive psychology is far more ambitious (Torrance & Pryor, 1995). Consistent with constructivist theory (Kamii, 1982) and the social constructivist theories of Vygotsky (1978; 1986) this type of formative assessment has been variously referred to as “divergent” (Torrance & Pryor, 1995), “strong” (Brown, 1996), “dynamic” (Dixon-Krauss, 1996; Glover & Thomas, 1999), “interactive” (Bell & Cowie, 1997; Cowie, 1997) and “educative” (Blackmore,

1988; Willis, 1994). This type of formative assessment is not focused upon identifying what students have achieved, but rather on what they might achieve and what they are ready to achieve with help (Torrance & Pryor, 1995).

Specifically, in this meaning of “formative”, assessment is integrated with learning, measuring a student’s assisted performance during collaboration with the teacher or a collaborating peer (Vygotsky, 1978, p. 86) (see also, Dixon-Krauss, 1996; Torrance & Pryor, 1995). Learning can then be “scaffolded” (Torrance & Pryor, 1995, p. 308) by the teacher providing appropriate tasks, questions and support. In essence, this type of formative assessment is characterised by

less detailed planning, open forms of recording (narrative, quotations, etc.), an analysis of the interaction of the child and the curriculum from the point of view of the child, and open questioning and tasks. (Torrance & Pryor, 1995, p. 316)

It is underpinned by a cognitive/constructivist view of education, an intention to teach in the zone of proximal development and a view of assessment as accomplished jointly by the teacher and the learner (Torrance & Pryor, 1995).

In 1995, Torrance and Pryor stated that exactly what the differences between a behaviourist and a constructivist approach to formative assessment might look like in practice were, at that time, “by no means clear” (Torrance & Pryor, 1995, p. 308). They reported that the evidence that existed on constructivist approaches to learning and assessment derived from very small-scale studies, often in non-formal education settings. Examples of such small-scale studies are research studies into the use of portfolio assessment in individual classrooms (see, for example, Crumpler, 1996; Fenner, 1995) and case studies of individual classrooms (for example, Borko & Elliott, 1998; Comber, 1996) or schools (for example, Anders & Richardson, 1994). However, since 1995, larger research projects have expanded knowledge about divergent assessment processes. For example, there have been focused observations in England infant classrooms (Pryor & Torrance, 1996; Torrance & Pryor, 1995) and, in New Zealand, research into formative assessment in science education by Bell and Cowie (Bell & Cowie, 1997; Cowie, 1997).

Torrance and Pryor videotaped over 100 hours of classroom interaction. They used a process of stimulated video recall, whereby the interviewer shows a video sequence to a participant teacher in its entirety, then replay it, pausing at appropriate points to allow the interviewee to comment and ask questions. Much of the assessment they saw in action seemed to be more of the convergent kind, focused on checking out learning outcomes. However, they did examine instances where assessment seemed more divergent, allowing the teacher insights into the learner's proficiency and scaffolding learning by intervening appropriately (Pryor & Torrance, 1996). In describing divergent assessment, Pryor and Torrance noted that although this view could be said to attend more closely to contemporary theories of learning, observed instances of such practice have not necessarily been found to be self-consciously "constructivist". "Rather they derive from more general ideological commitments to a 'child-centred approach'" (Pryor & Torrance, 1996, p. 208). Their findings suggested that a conceptual framework for understanding assessment that enables teachers to make use of the opportunities for formative assessment that they plan and that arise in their interaction with children was needed.

When constructing pupils' learning paths, teachers have to balance the creative and divergent views of the pupil with the need to structure learning experiences and pursue the convergent requirements of the curriculum. What seems to be required is an approach to assessment that enables both convergent and divergent teacher assessment to be pursued at appropriate times. This raises questions about how teachers might come self-consciously to perceive and manipulate different approaches to assessment, which might be the focus of more applied research. (p. 222)

In New Zealand, at Waikato University, Bell and Cowie (1997) used multiple data collection methods, including 128 classroom observations, 73 interviews with students and 65 with teachers, surveys, and audiotaping teacher development-day discussions, to investigate formative assessment in Years 7-9 classrooms in science in 1995. In addition, the teachers and the researchers met on 11 teacher development days for professional development in classroom-based assessment and to reflect on data analysis. Ten teachers from five schools

participated in the project. As a result the authors developed a model of formative assessment in science education similar to that of Torrance and Pryor (1995), in that two distinct, though complementary, processes were identified. Like convergent assessment (Pryor & Torrance, 1996; Torrance & Pryor, 1995)

planned assessment was that which the teacher had planned in advance to undertake in the lesson and it tended to be at the beginning of a lesson, for example, a question and answer session to ascertain what the students had learnt from the previous lesson. Planned formative assessment tended to be done with the whole class. It involved the teacher eliciting, interpreting and acting on the formative assessment information. (Bell & Cowie, 1997, p. 8)

In contrast, interactive formative assessment “occurred during the interactions between the teacher and the students, and so had the potential to occur anytime the teachers and the students interacted” (Bell & Cowie, 1997). This type of formative assessment involved the teacher noticing, recognising and responding to assessment information. This is very similar to divergent assessment (Pryor & Torrance, 1996; Torrance & Pryor, 1995). Interactive formative assessment tended to take place with individual students or small groups and could not be planned in advance. Harlen (1998), however, argued that teachers can be prepared for interactive/divergent assessment in advance. Bell and Cowie provided a model of how both of these kinds of formative assessment work in tandem in the classroom, linked through the purposes each serves. Their data demonstrated that some teachers use interactive assessment more than others and that teachers moved from planned to interactive and back depending on what they noticed as a lesson progressed.

This change was usually in response to focusing from the class to an individual. They usually switched back from interactive to formative in response to their responsibility for the whole-class learning. (Bell & Cowie, 1997, pp. 47-8)

Under stress, for example, when implementing new curriculum material, the teachers in Bell and Cowie’s study tended to use less interactive assessment. Additionally, a heavy emphasis on summative outcomes, for example, for the

ERO, also influenced the amount of interactive assessment they were able to accomplish (Bell & Cowie, 1997).

There are also links between this description of divergent assessment and “integrated codes” in terms of curriculum as explained by Bernstein (1971). Within an integrated type of curriculum, Bernstein explained the contents as standing in open relation to one another, and introduced the notion of “frame”, which “refers to the degree of control teacher and pupil possess over the selection, organisation and pacing of the knowledge transmitted and received in the pedagogical relationship” (1971, p. 50). Thus, in an integrated code, it would be possible for the teacher to blur the boundaries between subjects and use a more divergent approach to assessment. However, Bernstein specifically warns that one of the major difficulties which inhere in integrated codes arises over “what is to be assessed, and the form of assessment: the place of specific competencies in assessment” (1971, p. 65).

### Summative Assessment

In contrast to formative assessment, summative assessment is used in summing up and checking up on learning in order to report on a student’s achievement at a given stage of schooling (Broadfoot, 1995; Harlen et al., 1992). It usually takes place at the end of a unit, course or period of time and, due to the fact that it is usually reported, it is recorded in some way for this purpose (Gipps et al., 1995; Ministry of Education, 1994). Summative assessment results are often associated with a mark, grade or descriptive comment (Ministry of Education, 1994). Summative assessment can be of the “snap-shot” type, where, at a point in time, a teacher might decide to give the students an assessment in order to check up on where they have reached in a particular aspect or area of learning (Radnor & Shaw, 1995), or more formal, such as for reporting school achievement or awarding qualifications (Ministry of Education, 1994; Radnor, 1996). External accountability measures rely on summative assessments that provide the basis for comparisons between individuals and groups within schools, or between schools.

Summative assessment has a crucial role to play in providing information about the success of education generally and about progress and achievement for individuals and schools. This purpose of assessment is one of auditing learning for accountability (Darling-Hammond, 1997; Ministry of Education, 1993a; Wiggins, 1998). In recent years, many countries have introduced new systems of national testing that involve all students in the same year of schooling (Ministry of Education, 1998). But in each country that has sought to assess achievement nationally a different approach has been taken. For example, in England and Wales all 7-, 11- and 14-year-olds are tested and comparisons between schools have been made possible, while in Ontario, Canada, all 8–9-year-olds are tested in literacy and numeracy in order to set benchmarks for achievement at that age level (Ministry of Education, 1998). In the United States various states have implemented programmes ranging from standardised testing to state-wide portfolio assessments (Aschbacher, 1991; Herman, 1992). Across Australia standardisation is achieved through the provision of extensive illustrative material which teachers can use as a basis for judging their own students' performance (Broadfoot, 1995). But, in addition, the states of Victoria and New South Wales test the achievement of all students of a similar age and make the information available to schools and parents (Ministry of Education, 1998).

In New Zealand, since the restructuring by Tomorrow's Schools and the introduction of the new curriculum, national monitoring has been implemented through the National Educational Monitoring Project (Flockton, 1999). This aspect of New Zealand's education system "involves the systematic and regular collection, interpretation and reporting of information about important aspects of student achievement nationally" (Flockton, 1999, p. 3). Rather than national tests at the primary school level that assume that schools can be made to improve by offering students and teachers rewards for high test scores and penalties for low ones, this kind of measurement is more indicative of standards (Darling-Hammond, 1997). In Eisner's words, this function of summative assessment is "a kind of educational temperature taking" (1996, cited in Flockton, 1999, p. 3).



The key features of New Zealand's educational monitoring design operative over the years that the research reported in this thesis took place were as follows.

- National monitoring was administered at two levels: Year 4 (ages 8-9) and Year 8 (ages 12-13).
- Representative random 3 per cent samples of Year 4 and 8 students were selected each year (1440 students).
- The sample at each year level was divided into three groups of 480 students, comprising four students from each of the 120 schools. Three different sets of assessment tasks were administered to each of the groups. This allowed a considerable amount of information to be gathered without placing too many demands on individual students.
- Each year in a four-year cycle, covering all learning areas, two or three subject areas were assessed.
- The assessments were performance based and administered by experienced practising teachers selected nationally. These teacher administrators underwent specialist training for the task.
- The children's performances were videotaped where appropriate and all marking was centrally completed using experienced practitioners as well as trained markers.
- Reports on each subject and skill area tested were made available in the following year to government, schools and the general public. (cf. Flockton, 1999).
- The results provided a comprehensive national picture of the educational outcomes of nationally representative samples of students, including results by gender and by majority and some minority groups. (cf. Ministry of Education, 1998).

Another major use of summative assessment information is for making judgements about individuals and groups within each school in order to make them self-monitoring and self-managing (Blackmore, 1988; Filer, 1995), as described in Chapter Three. Born of the need to compare the achievement of students and

schools, this summative purpose, like national monitoring, also requires the objective measurement of results.

Most students, their parents and other stakeholders expect and often demand to know how well students are progressing at school in educational and other behavioural terms. Beyond the school itself major stakeholders such as the Government need to be able to consider investment and other decisions *in the light of actual school outcomes* and the quality of client satisfaction over time. (Education Review Office, 1995a, p. 1, emphasis added)

An emphasis on this purpose since the restructuring of schools and the curriculum in the 1990s in New Zealand has led to concern about the validity and reliability of school-based assessment (Education Review Office, 1995a; Filer, 1995). This

discourse of the technology of assessment is one in which the required *ends* – in this case comparisons of classroom results – are not in dispute. It is a discourse about the *means* whereby such ends can be achieved as fairly and objectively as possible. (Filer, 1995, p. 25, emphasis in original)

Worldwide, endeavours to ensure more reliable assessment outcomes have resulted in the development of various assessment instruments. Among these are the standard assessment tasks in Australia, the benchmarks project in Ontario, Canada, standardised achievement tests and systemising teacher assessment at Key Stages in England and Wales, the revised arrangements in Scotland, where all pupils between the ages of 5 and 14 take tests in reading, writing and mathematics when their teachers consider they are ready to, and the use of rubrics in portfolio assessment in parts of the United States. While New Zealand did not implement national tests for this summative purpose, a national set of assessment resources was established for school and teacher use during the period the field work for this thesis took place (Croft, 1999). These Assessment Resource Banks are collections of assessment tasks located on the internet. “They are organised to match the structure and terminology of New Zealand curriculum statements in mathematics, science and English, for levels 2 to 6” (Croft, 1999, p. 1).

Like the National Education Monitoring Project (NEMP), the Assessment Resource Banks (ARBs) were developed under contract to the Ministry of Education as a source of items that could be used by teachers as curriculum benchmarks. They were provided as complementary to schools' own assessment materials "to help teachers and schools to judge the relative performance of their students at given year levels" (Croft, 1999, p. 2). A strength of the ARBs is that they were prepared and trialled by teachers to assure acceptable levels of curriculum validity and test reliability (Croft, 1999).

Despite the introduction of the ARBs, however, schools throughout New Zealand have also had to invent their own mechanisms for this assessment purpose (Dixon, 1999; Faire et al., 1994; J. Hill, 1998a; M. 1998b). This was due to the fact that the ARBs have taken time to develop and schools have been audited and reviewed, both internally and by external agencies, since the introduction of Tomorrow's Schools in 1989. In 1995 the ERO reported that

few schools to date have recognised the need for assessment in relation to groups of students. Fewer still have been able to develop appropriate assessment procedures other than in some limited aspects of some curriculum areas for specific age groups. (1995a, p. 34)

The ERO went on to state that "aggregated information about student achievement is essential" (1995a, p. 34) for schools to meet the national assessment requirements and that teachers needed "considerable training in assessment practices and centrally developed assessment tools for their use" (p. 34). Furthermore, because comparisons need to be made across time to detect change, particular assessment regimes have to be kept in place long enough to achieve this.

A third use of summative information is for review, transfer and certification (Black, 1998). This information can be used within a school; for example, for communicating assessment information between teachers as students move to new classes or for reporting progress and achievement to parents and the students themselves. It can also be needed when students transfer to new schools. To be effective, the way the information is communicated to each audience is critical to its consequent use (Black, 1998; Faire et al., 1994). In the past, norm-

referenced grades were traditionally used for this assessment purpose (Black, 1998; Gipps, 1994a; Herman, 1992). Within an educative discourse of assessment, however, criterion-referenced methods of communicating this type of information have been developed. For example, Records of Achievement in England and Wales (Pole, 1993) and portfolio approaches in the USA (Crumpler, 1996; Darling-Hammond, 1997; Fenner, 1995; Tierney, Crumpler, Bond, & Bertelsen, 1996) and elsewhere (Paris & van Kraayenoord, 1992; Sutton, 1991). In conjunction with criterion-referenced tasks and portfolios, teachers' descriptive formative assessments have also been recorded and reported (Comber, 1996; Comber, no date; Faire et al., 1994; M.Hill, 1998b). This has been the predominant approach used in New Zealand in recent times. As described at the end of Chapter Two, Keeping School Records (Department of Education, 1989c), a system for communicating information about children's progress and achievement, was introduced into New Zealand schools in the 1980s. After many requests from schools in the 1990s, I was contracted by the Ministry of Education to conduct an action research project to trial draft formats of record cards and report forms in schools throughout New Zealand in 1997-98. A full report of this project is provided by M. Hill (1998a), but, in brief, the schools in the trial were very positive about having a standard format on which to record achievement and progress.<sup>4</sup>

### Interactions between Purposes

From the analysis above it is obvious that there are multiple purposes for assessment. Although I have chosen to use the formative/summative distinction to classify these purposes, multiple discourses run through and across these arbitrary boundaries. For example, within the conception of formative assessment, research

---

<sup>4</sup> The interviews of the teachers in the national sample for this thesis were carried out while I was undertaking the action research project for the Ministry of Education to trial record cards and report forms. Agreement was reached between the University and the Ministry of Education for this to occur. Chapter Seven contains further details.

has identified both convergent and divergent practices. In addition to multiple purposes for assessment there are multiple audiences for it, raising the possibility that one assessment task might provide information for several assessment purposes and audiences (Black, 1998). Harlen (1998), however, cautions that it is not always possible.

It seems only reasonable that assessment data collected for formative purposes, and so already available to the teacher, should be used when the teacher is assessing for summative purposes. Unfortunately there are significant disadvantages to this, depending on the way in which the term “assessment data” is interpreted. (Harlen, 1998, p. 9)

She explains that the main problem about this course of action is that because assessment for formative purposes is pupil-referenced and judgements are made about next learning steps, whilst summative assessment requires judgements to be made against public standards of criteria, external criteria tend to drive the process (Harlen, 1998, pp. 9-10). The higher stakes nature of the summative judgements, therefore, tend to overshadow the lower stakes status of informing learning. Harlen explained that she preferred to use the terms “assessment for formative and assessment for summative purposes” (1998, p. 1) because rather than have them appear as opposites, it is possible to envisage the procedures and purposes of assessment along a continuum. At one end, informal formative assessment purposes are met when teachers make pupil-referenced judgements during teaching to inform learning through feedback. At the other, formal summative standardised procedures are used with groups or individuals in order to award qualifications or report to parents and the community. Between these two ends of the continuum, the purposes and procedures can vary from more formative to more summative, depending on the purpose.

Other researchers have described the ways in which these and other purposes for assessment compete for attention (for example, Anders & Richardson, 1994; Bernstein, 1971; Broadfoot & Pollard, 1998; Crooks, 1988; Delandshere & Petrosky, 1994; Filer, 1995; Nisbet, 1993; Willis, 1994). Crooks (1988) drew attention to the critical importance of regular, relevant and specific

feedback to learners, and stressed that “most vital of all” (p. 470) assessment must emphasise what matters over what is manageable. Black and Wiliam extended these findings and reported a “clutch of problems [which] relate to the possible confusions and tensions, both for teachers and learners, between formative and summative purposes which their work might have to serve” (1998a, p. 59). These confusions and tensions result from a clash of what Foucault and others term “discourses” (Middleton, 1998; Weedon, 1987; Woods, Jeffrey, Troman, & Boyle, 1997). As Black explains, formative and summative discourses of assessment compete because

In so far as teachers have to be involved in both formative and summative assessment, they are bound to experience some tension between their two roles. The former is private, focused on the learning needs of their pupils, which it must be their first concern to serve. The other has to reckon with the pressures and constraints that come from outside. National and regional high-stake systems create pressures for teachers to work within a framework which drives both their school policies and parental expectations, and the personal concern for the best interests of their pupils might have to take second place. The teacher has to hold the boundary between the different requirements of the two roles. (Black, 1998, p. 120)

### **Competing Discourses of Assessment**

Formative and summative notions of assessment are not the only competing discourses that have been identified in the assessment literature. As argued in Chapter Three, effectiveness and accountability have emerged in the 1980s and 90s as important assessment discourses. Other researchers have also argued that there are competing discourses implicated in assessment and have identified a broader field of concern (for example, Bernstein, 1971; Filer, 1995; Woods et al., 1997). Filer (1995) contrasted a discourse of educational assessment, focused on improving teachers’ skills “regarding the achievement of objectivity and comparability of assessments through increased reliability and validity” (p. 25), with sociological discourses that raise issues about the use of

assessment for political and social functions. Some of the concerns of this alternative sociological discourse have centred around ways in which assessment can function to maintain the status quo, as an instrument of social control, and with the labelling of students.

Not least of the concerns of this alternative discourse has been the question of the fairness of grading, certification and selection processes. In this, sociology gives us some insight into the fact that assessment takes place within social contexts which systematically differentiate between social and cultural groups. Assessment results are shown to be a product of that differentiating process. (Filer, 1995, p. 25)

In a study about how teachers were assessed (appraised), Delandshere and Petrosky (1994) also raised issues about the critical nature of discourse in the role of interpretation. They argued that because an assessment system is always defined within an ideology, interpretations are framed within that ideology to produce consistent judgements. They contrast the ideology of essentialists (or positivists), who think of knowledge as truths or information that one possesses in a particular subject or field, with a Foucauldian post-structuralist perspective (Delandshere & Petrosky, 1994). In this view

knowledge is not a collection of discrete instances of truths; knowledge is rather, what people create, what they express, in discourses. Discourses create knowledge; they are not “signifying elements referring to contents or representations”.... Knowledge is also an object of discourse, created by individuals using language, but at the same time, individuals are themselves objects of discourse, creations of the language they use. Discourse is, Foucault argues, a system complete with its own *rules of formation and volitions* that have enormous control over what can be said within a particular field. Discourse is also “the group of statements that belong to a single system of formation,” and we are able “to speak of clinical discourse, economic discourse, the discourse of natural history, psychiatric discourse”...and, we would add, educational discourse. Like other fields, the field of educational measurement, for example, is created in discourse; it is also defined, for example, by statements that produce the notions of reliability and validity that define the boundaries of the field and control the possibilities of that particular discourse. Within a professional context, then, knowledge is produced in a discourse within a particular field,

and it is also constrained within the boundaries and possibilities of that particular discourse. (Delandshere & Petrosky, 1994, p. 12, emphasis in original)

This conception of knowledge allows for the idea that alternative discourses will continually be competing for teachers' attention. Bassey (1999), too, argued that the craft knowledge of teachers is influenced by a maelstrom of professional discourses. This is a maelstrom of ideas, theories, facts and judgements which the individual teacher meets (through informal discussion with other teachers, reading books and articles, participating in professional development and on the internet), broods on, contributes to and, occasionally, uses. However,

Within a discursive field...not all discourses will carry equal weight and power. Some will account for and justify the appropriateness of the status quo. Others will give rise to challenge to existing practices from within or will contest the very basis of current organization and the selective interests which it represents. Such discourses are likely to be marginal to existing practice and dismissed by the hegemonic system of meanings and practices as irrelevant or bad. (Weedon, 1987, p. 35)

Located in social institutions, such as schools and education systems, discourses are continually competing with each other for the allegiance of individual agents. The individual – in the case of this thesis, the teacher – is constantly subjected to multiple and conflicting discourses.

In thought, speech or writing individuals of necessity commit themselves to specific subject positions and embrace quite contradictory modes of subjectivity at different moments. It is the need to regulate disparate forms of subjectivity in the interests of existing power relations that motivates the language of common sense. Similarly, opposition to the status quo calls for attempts to unify certain subject positions and political interests. (Weedon, 1987, p. 97)

Sometimes the individual is aware that there is a choice. For example, during the conduct of the research for this thesis, there was a major debate taking place between the Ministry of Education and teachers, principals and education



academics, about whether a national system of externally-referenced tests should or should not be introduced at certain years of primary schooling (Eduvac, 1998; H. Lee & Lee, 1998; Ministry of Education, 1998; Phillips, 1998). This debate pitted the accountability discourses of assessment directly against the educative assessment discourses. In response to a discussion paper proposing externally-referenced tests, opinion was overwhelmingly against the use of such tests for accountability purposes (Gilmore, 1998).

At other times, the individual is unaware that there is a choice. For example, using the case of gender, Weedon wrote that

gendered subject positions are constituted in various ways by images of how one is expected to look and behave, by rules of behaviour to which one should conform, reinforced by approval or punishment, through particular definitions of pleasure which are offered as natural and imply ways of being a girl or woman and by the absence within particular discourses of any possibility of negotiating the nature of femininity and masculinity. (Weedon, 1987, p. 99)

Examples can also be generated from within assessment discourses. For instance, when the curriculum became constructed in levels with clusters of outcomes at each level, progress was assumed to be the achievement of ever more of the outcomes at increasingly higher levels. The action research project trialling record cards and report forms for the Ministry of Education in 1998, reviewed earlier in this chapter, found that, while it was not easy to make consistent judgements about performance at each level, the majority of schools and teachers believed it was possible to teach and measure progress in curriculum levels and used the record cards provided by the Ministry to do so (M. Hill, 1998a). It could be argued that by assuming it was possible to measure achievement and progress for all on this basis, teachers were unaware of any alternative ways of “monitoring progress against the achievement objectives” (Ministry of Education, 1993b, p. 3), and therefore could not choose to do otherwise.

Assessing individuals against a set of national criteria as both a way to improve learning and as the currency of accountability has become a popular way of attempting to increase school effectiveness in Anglophone countries in the past

ten years (Broadfoot, 1996). However, Nisbet (1993) stated that formative and summative purposes of assessment had proved almost irreconcilable in seven countries surveyed for a report commissioned by the OECD in 1991. In his speech to introduce the report in 1993 he outlined his fear that

In the attempt to reconcile these two very different requirements, the demand for accountability, for certification and selection,...will take precedence over the equally legitimate requirement that assessment should promote learning; that the two cannot be readily reconciled, and that, particularly in the design of national testing, accountability will take precedence over learning, because it involves "higher stakes". (Nisbet, 1993, p. 2)

Broadfoot and Pollard (1998) describe this re-formation of the curriculum as specified outcomes as the gradual establishment of a new hegemony of "performance". As already noted, a performance model "emphasises a specific output from the acquirer, a particular text that the learner is required to construct and the acquisition of the specialised skills necessary to the creation of the required output" (Broadfoot & Pollard, 1998, p. 14). In this way of thinking about teaching and learning, learners are aware of the learning outcomes that are valued and their performance is a means of locating them in terms of levels or some hierarchical judgement.

Willis (1994) has suggested that this performance model of criterion-referenced classroom assessment does not necessarily facilitate high quality learning due to the differing ideologies that underlie what she terms the technicist and educative views of assessment. In refining and extending Blackmore's (1988) description, Willis aligned the technicist discourse with that of using learning outcomes for measuring progress, and she stated that this "is a feature of market and management accountability" (p. 171). In contrast, educative assessment discourses reflect a more complex view of learning. In this view, learning is deemed to be a constructive process where individuals set out to create meaning.

As a result of the changes to education systems in the 1990s and the reactions of teachers to these changes, a growing amount of research has set out to investigate how the competing discourses and changes to curriculum and

assessment have impacted on students' achievement and teachers' assessment practices. Outside New Zealand, the emphasis appears to have been upon the effects of the introduction of national curricula, standards, authentic assessment and testing regimes. In New Zealand, however, as Chapter Two explained, there has been a national curriculum since 1877, and, for 60 years, national standards were either obligatory for schools or underpinned their practice informally. As set out in Chapter Three, the major changes for New Zealand schools in the 1990s were the fact that they became self-managing, with all that that entailed, and that the re-formed curriculum and assessment regulations required new knowledge and practices. The last section of this chapter reviews overseas research into the effects of these discourses on teachers' practice before focusing more sharply on recent New Zealand studies.

### **Overseas Research into Teachers' Assessment Practices**

The trend towards national standards, achievement targets and increased accountability throughout the OECD countries, and the consequent effects on schools, on teachers assessment practices and on student achievement, has attracted increasing research interest. Internationally, studies appear to be of two major kinds: the school effectiveness and improvement research, often, though not always, comprising large quantitative studies; and the usually smaller-scale, qualitative research into the effects of such changes on teachers' practice. As is clear from the arguments in this thesis, my interests lie in the second of these two fields of research and, thus, it is those studies that explore teachers' practice, and impacts on learners, that are examined here.

Although much research in assessment – even new approaches to assessment – has been dominated by technical discussions of validity, reliability and comparability (Torrance, 1998), studies of how teachers interpret and implement new requirements and approaches have been published in the 1990s. In England, for example, some relatively large studies have taken place (such as Broadfoot & Pollard, 1998; Gipps et al., 1995; Preece & Skinner, 1999),

accompanied by case studies carried out in particular schools (for instance, Filer, 1995; Radnor, 1996; Woods et al., 1997) or at particular levels of schooling (Torrance & Pryor, 1995). While some have used large-scale surveys and questionnaires supplemented by interviews (for example, Preece & Skinner, 1999), others have used interviews along with in-depth observations and other methods of data collection (for example, Broadfoot & Pollard, 1998; Gipps et al., 1995; Pryor & Torrance, 1996; Radnor, 1996; Torrance & Pryor, 1995).

Those focusing on primary school teachers have indicated that how school “reform” is interpreted, resisted and accommodated by the key actors – teachers – critically affects their assessment practice and therefore their ability to use assessment to improve learning (for example, Woods et al., 1997). My interest specifically in primary teachers’ assessment confined this review primarily to studies at that level of schooling. Relevance to the New Zealand context was also important in the selection of the studies reviewed here. Similarities were obvious, for example, in England and Wales, where a national curriculum and assessment system was designed not only to measure the performance of pupils, but also to make it possible for market forces to operate (Broadfoot & Pollard, 1998) – in other words, for schools to compete with each other for students. Several significant projects in England and Wales have investigated the effects, directly or indirectly, on primary teachers’ classroom assessment practice (Broadfoot & Pollard, 1998; Gipps et al., 1995; Gipps, McCallum, & Brown, 1996; Pollard, Broadfoot, Croll, Osborn, & Abbott, 1994; Pryor & Torrance, 1996; Torrance & Pryor, 1995; Woods et al., 1997) as well as at least one smaller-scale investigation (Filer, 1995). The focus of these investigations differed, however. Torrance and Pryor focused on teachers’ formative assessment practice; Gipps et al. investigated teacher assessment related to Key Stages 1 and 2; Pollard et al. on a longitudinal study of the impact of the changes on teachers’ practices in 48 schools; and Filer’s longitudinal study looked at the impact of the changes on one cohort of infant school children in just one school. Woods et al. (1997) drew information from four projects carried out after 1991 that focused on seven primary schools. Referring to the consequences of administration and curriculum restructuring, they

stated that the changes elicited complex and contradictory responses from the teachers.

Teachers filter the policies of change through their existing professional ideologies, perspectives and identities. This produces a variety of adaptations in the teacher workforce ranging from compliance with the new policy through mediation and accommodation to resistance and rejection. (Woods et al., 1997, p. 11)

Taken as a whole the studies in England in the early 1990s reveal a picture of assessment activity focused very much on performance goals. Through interviews and observations, all of these studies reported that teachers were becoming far more systematic in assessing, recording and reporting the achievement of their students. Not only were teachers more systematic in assessing performance, the Primary Assessment, Curriculum and Experience (PACE) project described how children, too, were “becoming ‘performance orientated’ rather than ‘learning orientated’ and the profound implications of this in influencing what is learned, when, for what purposes and in what way” (Broadfoot & Pollard, 1998, p. 10).

The children in the study were very aware of the importance of good marks, getting things right. In a climate of explicit and formal assessment, however, many of them avoided challenge and had a low tolerance of ambiguity. (Broadfoot & Pollard, 1998, p. 9)

Although these trends were more obvious at the Year 6 level in the PACE project, Pryor and Torrance, focusing more sharply on formative assessment in infant classes (Years 1 and 2), also reported that teachers were pervasively influenced by behaviourist ideas (1996). They reported the prominence of extrinsic rewards – stickers and the like – assertive discipline schemes and merit marks that counted towards school-wide competitions. They cited the work of Lepper et al. who argued that extrinsic reward systems have detrimental effects on intrinsic motivation.

Lepper’s work would suggest that the cultivation of intrinsic rewards, besides being important for enhancing student

motivation, also promotes more effective, deeper and longer lasting learning. (Pryor & Torrance, 1996, p. 211)

Even in formative assessment they reported a performance-goal climate whereby teachers do not grasp the divergent learning possibilities when children make mistakes. They are more likely to protect a child and pursue the convergent requirements of the curriculum. These authors reviewed at some length research to support the view that motivation to learn is bound up with the type of goals that students have. They cited studies in attribution theory and motivation to support the need to develop learning goals rather than performance goals in children in order for them to remain motivated to learn and to optimise their chances of success in learning. In particular, they cite Dweck's distinction thus

- a) *learning goals* in which individuals strive to increase their competence, to understand or master something new; and
- b) *performance goals* in which individuals strive either to document, or gain favourable judgements of their competence or to avoid negative judgements of their competence. (1989, cited in Pryor & Torrance, 1996, pp. 88-9, italics in original)

Pryor and Torrance summarised that Dweck found that children with performance goals

- avoid challenge when they have doubts about their ability compared with others;
- tend to self-handicapping so that they have an excuse for failure;
- tend to regard ability as a stable entity;
- concentrate much of their task analysis on gauging the difficulty of the task and calculating their chances of gaining favourable ability judgements;
- attribute difficulty to low ability;
- give up in the face of difficulty; and
- become upset when faced with difficulty or failure. (Pryor & Torrance, 1996, pp. 209-210)

She has also found that performance goals are developed in the early school years and become prevalent by the middle years of schooling, and that they are fostered by competition and encourage normative standards of success, i.e. comparison with peers. On the other hand, pupils with learning goals

- choose challenging tasks regardless of whether they think they have high or low ability relative to other children;
- optimise their chances of success;
- tend to have an “incremental theory of intelligence”;
- go more directly to generating possible strategies for mastering the task;
- attribute difficulty to unstable factors, eg. Insufficient effort, even if they perceive themselves as having low ability;
- persist; and
- remain relatively unaffected by failure in terms of self-esteem. (Pryor & Torrance, 1996, pp. 209-210)

It would seem, therefore, that teachers in Torrance and Pryor’s and the PACE (Broadfoot & Pollard, 1998; Pollard et al., 1994) studies, in order to raise standards, were employing curriculum and assessment strategies that engendered “the gradual establishment of a new hegemony of ‘performance’” (Broadfoot & Pollard, 1998, p. 11). In their quest to understand how policy to improve standards in education could lead to a situation that may in fact work against pupils becoming better learners, Broadfoot et al. have drawn extensively on the work of Basil Bernstein. They applied Bernstein’s conceptualisation of three message systems, “curriculum”, “pedagogy” and “assessment”, which he argued conveyed the existing “education codes” of their particular society, as a framework for gathering data in their research design. As a result of their early analyses the PACE researchers argued that “schools...teachers and pupils are embedded in a dynamic network of personal identity, values and understandings that are constantly developing in the light of internal and external interaction, pressure and constraint” (Broadfoot & Pollard, 1998, p. 11).

Because of this, policy directives are translated into classroom practice through a series of “mediations”. Such “mediations” we argued, should not be conceptualised in engineering terms as a series of articulated levers that relay a load through a structure. Rather they are the creative reinterpretation by the actors involved at each successive stage of the process of delivering education. (p. 11)

In summary, the PACE findings suggest that amongst teachers in the upper primary school there were increased feelings that priorities had been imposed on

the schools from outside, and a loss of fulfilment and autonomy. External accountability had increased and, although personal responsibility was still seen as important, there had been a shift from a covenant based on trust to a contract based on delivery of education to meet economic goals rather than as a form of personal development.

Some teachers expressed fragmented identities, torn between a discourse which emphasised technical and managerial skills, and values which continued to emphasise the importance of an emotional and affective dimension to teaching.... A new professionalism seemed to be emerging among newer teachers who were more likely (than more experienced ones) to find satisfaction within a more constrained and instrumental role without losing their commitment to the affective side of teaching. The issue of confidence was important to all teachers and lack of confidence to cover all the National Curriculum subjects may have contributed to the loss of fulfilment and enjoyment experienced by some teachers. (Broadfoot & Pollard, 1998, p. 12)

Filer (1995), too, in a longitudinal study of a cohort of children in their first years at school, raised questions concerning the limitations of what one can legitimately do with the assessments which varying classroom contexts give rise to. She argued that if assessments are to be used to indicate progress of successive standards, it is vital from the government's point of view that the message (the discourse) goes out that reliability can be achieved. Thus, in Britain, she argued,

one of the rhetorics of education promulgated by Her Majesty's Inspectors, the School Examinations and Assessment Council, Local Education Authorities and the educational press has been concerned with the possibility of "objective" and "systematic" teacher assessments as a realistic goal. (Filer, 1995, p. 24)

Filer highlighted two claims to objectivity associated with this assessment discourse. First, it is claimed that objectivity in teacher assessment can be achieved if teachers improve their skills of observing and assessing and reaching comparability of understanding of the meaning of the various achievement targets against which students are measured. This is a moderation argument. The second claim to objectivity is underpinned by the assumption that individual differences,



such as ability, effort, class or culture, are not relevant, but that recognition shall be given only to observable behaviour for which evidence exists. Filer's study demonstrated that there were fundamental differences between teachers that cannot be systematised out of existence in the pursuit of comparable contexts and objectivity in the assessment process. Their professional knowledge, both theoretically informed and grounded in experience, was in conflict with government inspired expectations that teachers be professionally equipped to make assessments which are comparable with those made in other contexts". In particular, Filer argued that

This conflict gives rise to further issues concerned with teacher accountability and I further argue the need for teachers to step outside government-imposed definitions of professionalism and accountability. I argue that there is a need for teachers to reassert their own professional knowledge and to disabuse parents and the wider public of the myth of "objectivity". (1995, pp. 35-36)

This recognition of fundamental differences between, for example, cultures or genders seems minimal in the assessment literature more generally. It should be noted, however, that in cross-cultural contexts it is not the case that one looks for different behaviours. The same observational skills might be used, but often the same behaviour may be understood differently; that is, it has a different meaning across cultures (personal communication with Ted Glynn).

Two other major studies, undertaken by Gipps and her associates, investigated the effects of the new National Curriculum and assessment requirements in England in the early 1990s. Both involved a sample of 32 schools. The first focused on teacher assessment (of students) practices at Key Stage 1 (7-year-olds) and the second investigated teacher assessment (of students) at Key Stage 2 (11-year-olds). In both investigations these researchers found that there was a range of assessment practices which varied along the dimensions of systematicity, integration with learning and ideological underpinning (Gipps et al., 1995; 1996).

The first study in 1992 pointed to three models of practice that were termed "intuitives", "evidence gatherers" and "systematic planners". Intuitives objected

to the new assessment methods as being a disruption to their intuitive ways of working. They relied on memory and did not have observable methods of ongoing teacher assessment, thus the researchers were unable to describe in detail the processes that this group was using to make assessments. Within this grouping, two subgroups were identified. “Children’s needs ideologists” showed a great deal of confidence and could articulate arguments about assessment that defended a child-centred view of curriculum, teaching and learning. “Tried and tested practitioners”, on the other hand, felt secure in modes of teaching and assessment practised before the introduction of the National Curriculum, but were less confident in articulating what these were, their basis, or how they used them.

“Evidence gatherers” believed that assessment must not interfere with teaching. Their main method of assessment relied on collecting evidence (such as work samples and artefacts) which they later evaluated. In this way they collected evidence that did not interfere with teaching practice.

Evidence gathering is associated with a belief that pupils generally learn what is taught and only what is taught; thus assessment follows teaching in order to check that the process is going to plan.... One of the main characteristics of evidence gatherers was that assessment was accommodated within existing systems and it is not always planned in. (Gipps et al., 1995, pp. 40-41)

“Systematic planners” planned time specifically for assessment, although this varied in the degree to which it was integrated with everyday teaching. The planned assessment of groups and individuals informed future task design and class work. Two subgroups were also identified within this group: “systematic assessors”, who gave regular concentrated time to one group of children at a time and had devised systems to lessen demands on them by the rest of the class; and “systematic integrators”, who did not separate themselves from the rest of the class, but circulated, gathering evidence in different ways which then fed into recorded assessment and informed planning.

Further up the school, at Key Stage 2, four distinct models emerged: “testers”, “frequent checkers”, “markers”, and “diagnostic trackers”. Testers went around listening and talking to children and keeping notes (some more frequently

than others). But this group did not feel that talking listening and note taking was sufficient. At the end of the taught units they gave planned assessment tasks, either self-designed or commercially produced. The tasks usually matched up with one or more statement of achievement. These teachers were very keen to have “proof” of attainment. Assessment for these teachers was at the end of each unit of teaching.

For frequent checkers the priority was teaching (rather than assessing), but teaching could be changed and adapted as a result of how children *in general* managed the activities from day to day or week to week. These teachers used the National Curriculum criteria and tasks were aimed at different statements of achievement. They did not particularly like the idea of formal testing. Data collection was an unobtrusive activity in most cases, with teachers noting things in the evening or at their desks, although one teacher went around with a visible checklist.

Markers made intuitive judgements using personal criteria and marking schemes that later needed to be converted into National Curriculum criteria before assigning a level. The work was aimed at the whole class, often at one particular level of the curriculum. Everyday classroom work in books, rather than special assessment tasks, was used as evidence and for assigning levels. These teachers were not interested in day-to-day data collection; they felt that taking observation notes or checking was obtrusive and took time away from teaching.

Diagnostic trackers planned for different National Curriculum levels, tracked children daily as they coped with their work and used questioning and observation, recording incidents as they happened. They aimed to integrate assessment with teaching, differentiating the work for children on different levels. They told children the criteria they were looking for in day-to-day work and made attempts to sit with individual children. Some set up times to review each child’s work. They kept evidence in the form of significant pieces of ongoing work to show development.

In comparing the Key Stage 1 (K.S.1) and Key Stage 2 (K.S.2) models, Gipps et al. saw intuitives (K.S.1) and markers (K.S.2) as similar, although the

markers relied on their marks as much as their memories. Testers (K.S.2), in that they planned assessment tasks after they had taught a particular unit, were similar to evidence gatherers (K.S.1). In this way, assessment did not interfere with teaching. But they also had elements in common with systematic planners (K.S.1) in that they needed “hard evidence” to back up their judgements. Frequent checkers (K.S.2) had definite similarities with systematic assessors (K.S.1) because both monitored children’s progress on the tasks they set. Diagnostic trackers (K.S.2) seemed similar to Systematic Integrators (K.S.1) because both groups integrated their assessment with their teaching, differentiating between students operating at different levels of the curriculum and providing relevant programmes to their needs. Gipps et al. state that they did not find teachers at the Key Stage 2 level who could be classified as Children’s Needs Ideologists. These teachers of younger children had “resisted criterion referencing as being in tension with ‘whole-child’ philosophy and were often confidently critical of the Standard Achievement Tasks as being inappropriate and ill-matched to their own ideas of ‘levelness’” (Gipps et al., 1996, p. 178). Diagnostic trackers (K.S.2) did carry the child-centred ideology of Children’s Needs Ideologists, but they did not resist criterion referencing.

An interesting finding from these studies was that teachers themselves were not easily able to articulate or analyse their own practice. They found it difficult to describe precisely what they did to determine levels of attainment or how they reached a decision. However, the researchers attempted to validate their models, at least in the Key Stage 1 study, by writing typical vignettes of practice and asking the teachers if each was like, or not like, them. The teachers responses provided a partial validation of the models the researchers described, in the sense that 31 (of the 32) teachers were prepared to commit themselves to one or other of the vignette models, thereby recognising and being able to identify with the models in practice.

Bachor and Anderson (1994), in British Columbia, Canada, identified groups of intuitives, evidence gatherers and systematic planners among 80 teachers at Grades 3 and 4, and Grades 6 and 7. Interestingly, the context in

Canada was one where teachers were being encouraged to move from grading to alternative assessment practices such as portfolio assessment and conferencing alongside a provincial examination system. In contrast to Gipps et al., who interviewed, observed, undertook quote-sort techniques and used vignettes, Bachor and Anderson used only interviews with the teachers in their study. In fact, a characteristic of the English studies reviewed here was that they all checked out their interview results through observing teachers in action. To some extent this could be due to the fact that several reported that teachers were fairly inarticulate about their assessment during interviews (Filer, 1995; Gipps et al., 1995; 1996), but Torrance and Pryor also maintained that it was necessary to use a methodology that enabled them to develop a full description of the nature and pattern of assessment activity (1995). Hence their use of audio- and video-tape recording of classroom action, allowing a full transcription as well as using these with teachers and students to re-visit assessment situations.

Some studies have used questionnaire-based surveys along with teacher interviews. Generally, these allow for a larger sample of teachers to be included. In Scotland, a larger study used a survey of 200 schools and 135 interviews to investigate the effects of curriculum and assessment changes there (Harlen, Chisnall, & Byrne, 1995). Rather than testing all children at Key Stages as in England, in Scotland national tests were developed at each level of the curriculum. These were to be given by teachers when they thought the children were achieving at a particular level as a means of checking up on their own assessments. In this way they acted as a moderation device. Investigating teachers' assessment practices after the introduction of these tasks using questionnaires and interviews, Harlen et al. (1995) reported that teachers were becoming more systematic and wide-ranging and recording more. However, they found that the majority of the teachers had not taken on board the moderating role of the test material. They also stated that these teachers found it difficult to describe the deeply embedded aspects of their practice or theories of learning that underpinned them. These teachers often saw recording of assessment information as an end in itself and tended to look for "hard" evidence.

In contrast to the relative uniformity of the studies and their findings in Great Britain, studies in the United States are far more difficult to classify. To a great extent this is probably due to the extreme variation in education, curriculum and assessment policies across states, there being no national curriculum at all. In fact Hargreaves (1999) contrasts North American responses to the uncertainties of the postmodern world with the conservative centralising and managerial approaches that may have comforted governments in places like Britain and New Zealand with procedural illusions of effectiveness. He believes that in North America governments “have rolled with and embraced...uncertainties and complexities, valuing multiple intelligences, diverse learning (and teaching) styles, and a process-based rather than a content-based curriculum” (Hargreaves, 1999, p. 15).

Darling-Hammond identifies two major approaches to raising standards in the USA. Some, she argued, have put their faith in state standards and tests, believing that they would provide a trigger for rewards and sanctions for students, schools and districts that would motivate students and teachers and drive reform (Darling-Hammond, 1997). In the United States particularly, standardised tests have been a powerful force in shaping classroom life (Darling-Hammond & Wise, 1985), though not necessarily in a positive direction. In a study of 43 teachers Darling-Hammond and Wise reported that most commonly standardised testing altered the curriculum emphasis, led to teaching students to take tests, encouraged teaching to the test, reduced time to teach (and learn), and promoted a feeling of being under pressure (1985, p. 349). Johnston (1989) extended these findings, adding that assessment actually altered how reading, for example, was taught, as well as what was taught. Other U.S. researchers have also made it quite clear that standardised testing has “run amok” (see, for example, Anders & Richardson, 1994; Firestone & Winter, 1998; Shepard, 1989). Test-based accountability schemes have had similar effects. As Darling-Hammond argued

Focusing on testing without investing in organizational learning is rather like taking a patient’s temperature over and over again without taking the necessary steps to promote greater health. (Darling-Hammond, 1997, p. 241)

Instead of this carrot-and-stick approach, some states attempted to use more authentic assessment approaches. “Authentic assessment” is a generic term that is used to describe a range of approaches to assessment that are alternatives to standardised testing. The implication of the term is that assessment tasks should be more practical, realistic and challenging than more traditional “pencil and paper” tests (Torrance, 1995). Kentucky and Vermont both set out to create more performance-based assessment systems in the late 1980s. However, they used different approaches. Kentucky’s system was framed by an accountability structure that proposed rewards for schools whose average scores (presumably on rubrics) improved year by year and sanctions for those who did not improve to the level specified in the statute. Vermont’s system was launched with the expressed aim of focusing attention upon how instructional improvement might be achieved. Indications are that Vermont’s approach has encouraged the wide involvement of teachers and proved a powerful tool for improving teaching (Murnane in Darling-Hammond, 1997, p. 242). This is supported by the results of other studies about the impact of portfolio and performance assessment on teachers. Specifically, results have suggested that these assessment methods can lead to more child-centred classrooms, encourage higher-order questioning, encourage a more integrated curriculum and involve parents more closely (Crumpler, 1996; Fenner, 1995; Viechnicki, Barbour, Shaklee, Rohrer, & Ambrose, 1993).

Anders and Richardson (1994) however, demonstrated that where the context has been a standardised testing one, it might take time and much encouragement for teachers to trust their own assessment judgements and learn to develop their own systems. On the East Coast, a think-aloud method for investigating teacher report writing demonstrated how complex tensions and conflicting values affected teachers during this process (Afflerbach & Johnston, 1993). This study explored how 11 teachers from three school districts, each with different levels of control over their curriculum and grading practices, resolved tensions and conflicts. The findings exposed an unanticipated variety of conflicts, the stressfulness of the process on the teachers, and an unanticipated level of constraints under which the teachers worked. One or two quotations from the

teachers' think-aloud transcripts give an impression of how competing discourses and multiple audiences for reports impacted on these teachers.

This little girl has two incompletes, which means she was there and that she did not finish her work...and she gets an F...five grades divided by seven...not very high...so she has a 58...but I'm already covered...when I put it on her report card...I don't like to do that...but I have to...even if there is a borderline I can't change the grade...I can't help the grade...for one thing...our district manual...says...these numbers are what I have to use to be accountable. (Afflerbach & Johnston, 1993, p. 78)

The whole process is frustrating...see...in graduate school...and I liked the program I was in...we looked at how to use portfolios in the classroom...that seemed to be like a very good way...a comprehensive way to look at what a student does...so I keep a portfolio...for each student...but it hardly matters.... I mean it matters to me.... They want to move teachers out of the schools that don't improve by a certain percentage on the standardised tests...and the report card has information related to the tests...so it's not only that the portfolios I have don't figure much...it's also that the alternative, their chosen alternative, is a bunch of test scores. (Afflerbach & Johnston, 1993, p. 81)

Australia, too, has different systems for raising standards in each state, though national benchmarks have been provided as guidelines. Comber (1996), investigating literacy teaching and learning in a disadvantaged school in South Australia, used documentary analysis, observation and, by treating oral data as text, Foucauldian analysis techniques in order to "make the familiar strange" and consider the effects of discursive practices. Like Afflerbach and Johnston (1993), Comber also used the medium of teachers' reports on children's literacy to explore discursive tensions. She investigated students' school reports written over the years in which economic rationalism had become a distinct influence on the curriculum and assessment requirements in Australia. She was able to identify significant changes in how students and learning were constructed by teachers. Quoting Donald she stated

records kept on pupils in the poorhouse and industrial English schools of the nineteenth century were "modelled on the files kept by the police and the prisons". School reports were



developed out of a concern with monitoring and regulation of populations (including poor children) who might otherwise be dangerous. The normalising function and disciplinary power of reports may at first be more difficult to detect in contemporary progressive and managerial discourses to which Carlo is subjected, but as Dyson demonstrates these professional conceptual tools for envisioning child literacy development “write the child” through “grids of identities” (following Foucault 1970) which delimit the pedagogical relationship and ultimately whose children “can be in the community”. (Comber, no date, p. 27)

Comber probed the extent to which children’s actual identities and life worlds are linked to their success and failure at school. She demonstrated how these successes and failures are inextricably linked with teachers’ discursive practices and to technologies of assessment. She argued therefore that

For these reasons it seems important for educators across educational sectors to begin to interrogate how our own discursive practices in everyday institutionalised assessments exercise power in ways which limit who we can be as teachers and constrain the kinds of literacies we make available to students. (Comber, no date, p. 27)

Another Australian study of assessment that used a Foucauldian genealogical approach was an investigation that sought to understand how assessment and selection in education have come to be viewed in particular ways (Meadmore, 1993). Meadmore explicated a range of “dividing practices” including the State scholarship examination, several psychological testing programmes, the “October tests” and the current competency-based performance standards. Her thesis demonstrated how these practices were technologies deployed for reasons of governmentality. They were also shown to produce “scholastic identity” on an individual basis. Similarities between the situation investigated by Meadmore in Queensland in 1993 and in New Zealand schools around 1996 suggested that using a similar theoretical framework may assist me to investigate the effect of assessment practices on teachers in self-managing schools in New Zealand.

## **Research on Teachers' Assessment Practices in New Zealand in the Early 1990s**

In New Zealand, soon after schools became self-managing in 1989, the New Zealand Curriculum Framework (NZCF), which set out the foundation policy for teaching, learning and assessment in New Zealand schools, was published. Although clear differences existed between this policy and curriculum and assessment policy in England and Wales at the time, there were also some “striking similarities” (Sutton, 1994, p. 334). The framework set out seven essential learning areas (language and languages, mathematics, science, technology, social science, the arts, and health and physical wellbeing), each subdivided into sets of specific objectives (known as “achievement objectives”). The radical change from past national curriculum documents in New Zealand was the production of these achievement objectives. These were set out in a number of progressive levels (usually eight), with assessment procedures designed to provide information on student achievement against these objectives. The information recorded from the assessments was then to be used to build a profile of individual achievement (Aikin, 1994).

It is this model of “progression” determined by assessment achievement on a linear and hierarchical route that appears to be in conflict with the “child-centred” approach espoused by primary teachers in New Zealand. (Aikin, 1994, p. 58)

Although national tests such as those in England and Wales were also debated, they were not introduced into the New Zealand system. Instead, the National Education Monitoring Project (Flockton, 1999), introduced earlier in this chapter, was adopted. However, the notion of individual achievement through meeting objectives at progressively higher levels in the curriculum is strikingly similar to the situation in Great Britain and Australia. Such an approach presupposes a clear view of what constitutes progression in each curriculum area, and a “one size fits all” model view of learning. Aikin (1994) raised concern that the new curriculum and assessment requirements had the potential to undermine child-centred teaching practices, and cause teachers to become more summative in

their assessment practices so that results might be aggregated. Wylie (1994), too, put this succinctly.

The increased accountability of individual schools in the current forms of accounting for public money puts emphasis on demonstrating results, and hence encourages the use of aggregated children's individual achievement data to provide such evidence. (Wylie, 1994, p. 120)

Aikin (1994) stated that it was, at that time, difficult to judge precisely how primary teachers, working mainly as generalists and teaching all the required areas of the curriculum, would be able to teach, record and report on the progress of each child against the relevant achievement objectives. She observed that there were signs that the situation had already led to a "massive increase in workload for primary teachers keen to do their best for each student" (Aikin, 1994, p. 70). Other researchers and commentators, too, noted this trend (Calder, 1995; Livingstone, 1994), including Ramsay et al., who termed the situation at that time "an assessment frenzy" (1995, p. 63). Research into these changes on teachers' assessment practices began to flourish.

With the introduction of Keeping School Records in 1989, criterion referencing became the norm in New Zealand primary schools (Faire et al., 1994). A national workshop on professional development in educational assessment, held in Wellington in 1992, concurred that

The main purpose of assessment is improved student learning, [and] a basic philosophic shift is occurring in New Zealand from norm-referenced assessment procedures towards standards based and criterion-referenced assessment. (Ministry of Education, 1992, p. 150)

An exploratory study carried out in the Waikato and Bay of Plenty in 1993-94 confirmed that even before the National Education Goals and the National Administration Guidelines (NEGs and NAGs) (Ministry of Education, 1993b) were implemented, teachers in primary schools were using criterion-referenced assessment systems

- based on comment on individual achievement rather than on comparisons between children;

- the comments are related to learning goals rather than comparative ratings of what children can do;
- each teacher is required to keep a cumulative record of specific information on each child's progress in learning, and in personal and social development. (Faire & Yates, 1994, p. 115)

The same study found that most of the information recorded on the record cards was sourced from the school's own systems of individual records (Faire et al., 1994). Many of these comprised files kept on each individual student, stored in clear plastic envelopes, manilla folders or an exercise book in which data was written and work samples were pasted. The teachers in the study explained that they added continuously to these files, "building up a rich data-base about learning gains, goals, styles and needs" (Faire et al., 1994, p. 9). These systems were designed and used as working documents rather than storage systems of results of progress. Some schools and teachers involved the students in the compilation of these files, while others were completed entirely by the teachers. Other records were also kept by teachers, including: checklists; graphs of progress, particularly in reading, oral language and mathematics; running records of reading; and anecdotal notes about the observations they had made (Faire et al., 1994, p. 9).

In order to keep such records and inform their teaching practice, primary teachers gathered assessment information using a range of strategies that have been reported under the headings of

- observation;
- conversing, conferencing, and interviewing;
- collecting and analysing work samples;
- keeping anecdotal notes;
- running records of particular event, particularly reading running records;
- taping events for later analysis;
- testing;
- involving students in self/peer assessment;
- discussions with other teachers and parents. (Faire & Yates, 1994; Kirk & Gilmore, 1995).

These strategies are nearly all examples of what is referred to in the educational assessment literature as “performance” or “authentic” assessment (see, for example, Gardner, 1992; Gipps, 1994a; Linn, 1993; Stiggins, 1997). Faire and Yates (1994) reported that many types of tests, including multiple choice and standardised tests (for example, the Progressive Achievement Tests), had lost their popularity with primary teachers. The kinds of tests mainly reported as being used in primary classrooms tended to be of the performance variety: teacher designed pre-test/post-tests in mathematics, and skills testing in physical education (Faire et al., 1994, p. 119).

A survey of school entry assessment practices (Thackery, Syme, & Hendry, 1992), conducted for the Ministry of Education, randomly surveyed 100 schools nationally about the practices they used to carry out assessment of 5-year-olds as they entered schooling. The investigators also interviewed teachers in 26 schools throughout central New Zealand to explore these practices in greater depth. Revealing a considerable degree of uniformity, the teachers in the sample collected comprehensive and useful information to assist teaching. Anecdotal observation was noted as by far the most frequently used technique but these teachers also adapted diagnostic tools such as the Concepts about Print and the Sand and Stones tests (Clay, 1979), and these instruments “were widely used and understood by teachers” (Thackery et al., 1992, p. 39).

However, by 1995, following the implementation of the first new curriculum documents and the NEGs and NAGs in 1993, primary teachers seemed to be losing confidence in their ability to keep pace with new assessment requirements. Faire et al. (1994) reported that the overall opinion of the teachers in their study was that the school record cards were inadequate and not meeting the schools’ needs; that there was a lack of consistency of information; that there was a need to take accountability purposes into consideration to demonstrate that they were meeting the official requirements; that there should be record cards that reflected the nature of the revised curriculum; that data should be recorded electronically; and that there was a need for systems particularly tailored to the needs of Kura Kaupapa Māori (Māori immersion schools).

Milestone reports from professional development contracts confirmed that primary teachers needed help to extend their knowledge about assessment ideas, especially as these related to the NZCF (Ministry of Education, 1993a), the development of effective and efficient assessment procedures for each school, the design of new school assessment policies, and the design of reasonably valid and reliable assessment tasks for use school-wide (Calder, 1996). Furthermore, research in the Waikato in 1995 demonstrated that at least nine changes in teachers' assessment practices could be identified since the introduction of the restructured education system in early 1990s (Calder, 1995). The four most significant changes among the nine were identified as

- changes arising from the implementation of the new curriculum;
- a change towards more individual assessment;
- changes to meet the needs of the Education Review Office;
- more emphasis on record keeping. (Calder, 1995)

Wylie (1994), in the fifth year of a longitudinal research project surveying 239 primary school teachers about self-managing schools, noted a 23 per cent increase from the previous year in teachers reporting substantial change to the curriculum they taught. Her results indicated that there had been a definite increase in the amount of assessment since the shift to school-based management – 74 per cent of teachers reported that the amount they had done had increased over the last year. While 49 per cent of teachers reported that this had improved their view of individual learners, only 14 per cent had made changes to their curriculum as a result. On the down side, 35 per cent of teachers reported less time to give individual attention and 25 per cent said there was less time to “cover the curriculum” (Wylie, 1994, p. 118).

Wylie's results also revealed that although teachers reported doing far more assessment, there was little change in the kinds of assessment they reported doing. These procedures were very similar to those reported by Faire et al. (1994). About half the teachers surveyed reported some aggregation of individual children's results although, in contrast, only 17 per cent of principals surveyed reported no aggregation occurring in their school (Wylie, 1994). Wylie

summarised the results, stating that they showed “scant sign of radical change in curriculum and assessment” (Wylie, 1994, p. 121).

But the survey findings also show that the new environment in which schools operate is having some affect, perhaps more subtly than directly. More attention has been paid to assessment and reporting to parents at the classroom level. ...The integration of assessment results with school reporting and planning systems is yet to occur on a widespread basis. (Wylie, 1994, p. 121)

At about the time that the investigation reported in this thesis began, in 1995, several complementary investigations were also initiated. Results from these are reported here because I was in touch with the authors both before I undertook this study and throughout the time this study took place. As I explain below and in Chapter 5, the methodologies used in these projects, and the findings from some, had an influence on the decisions I made in my investigation.

Three recent studies in New Zealand have combined research with action research and professional development contracts. Bourke and Willis (1998) investigated how primary teachers grappled with tensions between assessment for accountability and assessment for formative purposes in the context of an Assessment for Better Learning contract for the Ministry of Education. Using interviews and observations in one case study school, Bourke and Willis reported that after the professional development programme the teachers were more confident about classroom assessment and the assessment policies within the school, and that they were more open to new approaches to assessment. They also reported that the teachers’ views about how learning took place tended to affect their practice. They argued that in self-managing schools the nature of change was idiosyncratic and that teachers and principals often held different views about the nature of learning, which had an effect on assessment systems because the way teachers viewed learning was inherent in the assessment practices they chose. In other words, they were operating according to a range of different discourses. Thus, within a school, there could be a clash of views and discourses of both learning and assessment. In attempting to resolve these tensions, the professional development programme worked to dispel the notion that there was “one right

way” to implement school-wide assessment. Bourke and Willis quoted Fullan (1998) in “realising that there is no answer [which] can be quite liberating” (Fullan, 1998, p. 10).

Jan Hill (1998a) also investigated the effects of change on practice in an Assessment for Better Learning professional development contract for the Ministry of Education. She used an action research case study approach in four schools in the Auckland region. Group discussions, individual interviews with teachers, observation, questionnaires, documentary analysis and an assessment help desk were all used to gather data for this investigation. Her findings indicated that although the professional development appeared to move teachers in a formative direction, summative assessment seemed to dominate formative in classrooms. She attributed this to a lack of teacher knowledge and confidence about assessment in an environment where teachers felt more accountable. She reported that schools appeared to operate much as educational islands, with reasonably consistent records about individual student progress shared within each school but a lack of consistency in the information that was shared between schools.

It was not clear to what extent the teachers in this study were observed to confirm that their practice was consistent with the information gained from interviews, nor how frequently they were observed nor how the data were analysed. However, tensions between formative and summative assessment similar to those reported in the overseas studies and by Bell and Cowie (1997) and Bourke and Willis (1998) were described (J. Hill, 1998a).

Dixon, too, studied teachers’ assessment practices in four Auckland schools in 1997. Two of the schools had recently had ERO reviews in 1996, while the other two had not and were not expecting reviews in 1997. Of the schools in each category, one was large (over 20 teachers), the other had fewer than ten teachers. A comprehensive questionnaire about assessment practices, beliefs and attitudes was administered to 43 teachers in these schools, and 12 teachers (three from each school) were then interviewed (Dixon, 1999).



In all of the four schools in Dixon's study assessment had been a critical issue for teachers and had been a focus for school-wide professional development. The teachers in all four schools had been working together collaboratively, often in syndicates, to effect and manage curriculum and assessment change. However, their attempts to define what constituted good assessment practice appeared to have been tempered somewhat by the perceived expectations of external agencies such as the ERO (Dixon, 1999). For example, although they valued formative assessment highly, they had increasingly implemented new summative practices for accountability purposes. Specifically, Dixon reported that although the teachers in these four schools attributed greater importance to formative assessment for its potential to enhance learning, their practice was dominated by summative activity. The findings attributed this to both the teachers' lack of understanding of the critical aspects that make assessment formative, and the perceived need to meet the external accountability demands. Often they perceived these expectations as being in conflict with their personal beliefs about teaching and learning, but they reported that assessment for accountability still monopolised their time and energy.

### **Summary and Comment**

This chapter reviewed the contemporary literature about classroom assessment, particularly formative and summative assessment, the issue of accountability and the maelstrom of discourses of assessment with which teachers in primary schools have to deal. It spotlighted international research carried out since the major educational restructuring of the 1980s and highlighted investigations into the effects of the structural changes to school administration and curriculum and assessment on New Zealand primary teachers in the 1990s.

The literature reviewed in this chapter indicated that the field of contemporary assessment research is rapidly expanding, complex, and, at times, confusing. For instance, the description of assessment types is sometimes confusing. Different practices are described differently by the same authors at

different levels of schooling, (for example, Gipps et al., 1995). Apart from studies of formative assessment (for example, Bell & Cowie, 1997; Torrance & Pryor, 1995), there appears to be almost no evidence of evolving assessment practices having been utilised by teachers to teach any differently. And, except for Filer's (1995) and Comber's (1996; 1997; no date) work, it is interesting how little aspects such as culture, gender and social class feature in the assessment research literature.

As I have indicated in Chapter One, in my role as a teacher educator and researcher I was very well aware of the assessment frenzy identified by Ramsay et al., (1995) and was thus interested to investigate for myself the effects of educational and managerial assessment discourses in actual primary classrooms. Specifically, my interest was in the ways in which managerial and educational discourses were impacting on what primary school teachers could do and say. These discourses gave rise to a number of questions for me to grapple with as I tried to assist teachers acquire new knowledge and deal with policy requirements. What effects do repeated claims about failing schools have on teachers in classrooms? How does a new urgency and anxiety about standards, as a matter of government concern, intersect with professional educational discourses about assessment and pedagogy? To what extent does the proliferation of discourses about assessment result in a re-disciplining of teachers and students? To what extent do the curriculum and assessment requirements for monitoring students' learning discipline teachers' educational beliefs and practices? How are these discourses taken up, used and resisted by teachers? And how do these discourses of education and schooling impact upon and discipline my own practices as a teacher educator?

In order to answer these questions it was necessary to investigate real classrooms and the assessment practices of teachers. Influenced by Foucault's concept of discipline, which encompasses everyday routines (e.g. checking off outcomes), classifications of time (e.g. timetables and schemes of work), and compartmentalisations of space (e.g. the arrangement of groups or individuals within classrooms), I deemed it necessary to observe and question teachers in the

conduct of their everyday practice. Chapter Five describes how I went about exploring and excavating teachers' assessment practices and analysing the effects these competing discourses had upon their practices.

## **CHAPTER FIVE**

### **RESEARCHING TEACHERS' PRACTICE**

Educational research is critical enquiry aimed at informing educational judgements and decisions in order to improve educational action. This is the kind of value-laden research that should have immediate relevance to teachers and policy-makers, and is itself educational because of its stated intention to "inform". It is the kind of research in education which is carried out by educationalists. (Bassey, 1999, p. 39)

For a long time ordinary individuality – the everyday individuality of everybody – remained below the threshold of description. To be looked at, observed, described in detail, followed from day to day by an uninterrupted writing was a privilege. The chronicle of the man, the account of his life, his historiography, written as he lived out his life, formed part of the rituals of his power. The disciplinary methods reversed this relation, lowered the threshold of describable individuality and made of this description a means of control and a method of domination. It is no longer a monument for future memory, but a document for possible use. (Foucault, 1977, p. 191)

In early 1996 I was thrust into the position of having to make a decision about how to explore the ways that teachers in self-managing schools were responding to the competing discourses of assessment examined in Chapters Two, Three and Four. Late in 1995 my proposal had been to collect information from a representative sample of teachers. It was envisaged that a questionnaire would be used for this, followed by open-ended interviews to delve into the findings in more depth. A second phase would then involve working in partnership with a small sample of teachers to help them describe and analyse their assessment situations in depth, plan with them appropriate action to address the situation, collaborate on the implementation of the plan, observe and record the effects of that action and reflect together on the outcome(s) of these changes. This process would be carried out several times, in a cyclical process of action research (W. Carr & Kemmis, 1986; Cohen & Manion, 1989; Elliott, 1991; Kemmis & McTaggart, 1988). My

intentions were based on my belief that educational research is critical enquiry aimed at informing educational judgements and decisions in order to improve educational action (Bassey, 1999; Zuber-Skerritt, 1995).

Things did not go exactly to plan, however. In February 1996 I was contacted by the principal of a primary school and asked to assist his staff in reviewing and improving their assessment systems and practices. As I have explained in Chapter Four, by 1996 there were research results emerging that demonstrated that primary teachers in New Zealand were struggling to reconcile their assessment practices with accountability demands (for example, Calder, 1995; Faire et al., 1994; Wylie, 1994). Therefore, on the basis of this literature and after discussions with the teaching staff at the primary school, I decided to proceed directly to working in the school. Shortly after I had made this decision, another school approached me with a similar request. This chapter outlines the research process I utilised and my reasons for it.

The first part of the chapter addresses the multiple and competing discourses of research and the way I have attempted to navigate my way through these. It also presents the rationale for selecting an interpretive approach and my reasons for employing a mixed method approach. The second half of the chapter outlines my understanding and use of a participative case study approach and the research questions and methods I used to gather information about teachers' assessment practices.

## **Multiple Discourses of Educational Research**

Thirty years ago Schwab urged that educational research pursuits be redirected from theoretical pursuits "such as the pursuit of global principles, comprehensive patterns, the search for stable consequences and invariant elements, the construction of taxonomies of supposedly fixed or recurrent kinds" (1969, p. 288) to studies of the practical. Beattie (1997) has argued that not much has changed and that a great deal of the educational research in the 1990s is concerned with questions that are topical in the social and behavioural sciences

rather than with issues that confront practitioners in schools. She pointed out that this kind of logico-scientific research is focused on the development of theory, uses methods, goals and questions that originate in the academy rather than in schools, and is presented in language that is propositional rather than narrative, remote from practitioners' ways of knowing and acting, and "does not resonate with the way they understand the realities of their lives" (Beattie, 1997, p. 4).

But much has changed in educational research. The process-product research of the 1970s raised issues about research design. In this paradigm, teachers' thinking and teachers' practice were regarded as if they were separate but related. They were explained through psychological lenses that failed to account for the complexities of teaching or the ways in which teachers know their work.

Researchers came to understand how traditional approaches to educational research conducted in the empirical/analytical research paradigm, which employed quantitative and statistical methods, experimental designs, correlations and variables, left out too much that is of value when they are employed in the study of humans. These research methods whose purpose is to make context-free generalisations and to focus on the key issues of reliability, validity and replication, have their origins in scientific research and were generated for the study of the sciences and the natural world. (Beattie, 1997, pp. 15-16)

Studies such as these are important because their results have indicated in general where learning gains are made and, for example, how evaluation strategies such as tests and grading impact on student learning (Crooks, 1988; Madaus, 1988). Black and Wiliam (1998) reviewed at least 20 such studies that included innovations that strengthened the practice of formative assessment and led to "significant, and often substantial, learning gains" (p. 3). Their statement that there is such a firm body of evidence that formative assessment is an essential feature of classroom work, that development of it can raise standards and that they know of no other way of raising standards for which such a strong *prima facie* case can be made is strong indeed.

But as Black and Wiliam (1998a) themselves point out, although it might be seen as desirable to attempt a meta-analysis of the quantitative studies that have been reported to identify causative factors for improvement, the fact that it hardly

seems possible prompts a reflection on this field of research. The reports they studied brought out other important contextual features critical to the use of such assessment for increased learning. The crucial aspect of each of these was that it is what teachers and students believe about learning, and how they behave in their teaching practice, that makes assessment formative or not.

Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles which leave entirely to them the task of translating them into everyday practice – their classroom lives are too busy and too fragile for this to be possible for all but an outstanding few. What they need is a variety of living examples of implementation by teachers with whom they can identify and from whom they can both derive conviction and confidence that they can do better, and see concrete examples of what doing better means in practice. (Black & Wiliam, 1998b, pp. 15-16)

The problem with traditional educational research conducted in the empirical/analytical research paradigm into formative assessment, for example, is that it fails to give clear accounts on one or other of the important details (Beattie, 1997; Black & Wiliam, 1998b) – for instance, about the actual classroom methods used, or about the motivation and experience of the teachers, or about the nature of the tests used as a measure of success, or about the outlooks and expectations of the students involved (Black & Wiliam, 1998b).

Dissatisfaction with these methods gave rise to a search for more suitable, qualitative methods which could take researchers into classrooms and the lived experience of teachers' lives. Sherman and Webb (1988) explained that this type of research implies a direct concern with experience, as it is "lived" or "felt" or "undergone" (p. 7). Qualitative data enabled researchers to interrogate the dilemmas, understandings, feelings, values and experiences of the teachers in schools as they occurred. Qualitative researchers sought attachment, or what Polanyi (1958, cited in Beattie, 1997) calls "indwelling", and ways to inquire into the meanings, values, moral purposes and intentions of those who work in schools.

Many writers see the differences between quantitative and qualitative paradigms as a dichotomy (for example, Cohen & Manion, 1989; Patton, 1990; Smith & Heshusius, 1986). On one side, logico-postivism uses quantitative and

experimental methods to test hypothetico-deductive generalisations, while, on the other, phenomenological inquiry uses qualitative and naturalistic approaches to inductively and holistically understand human experience in context-specific settings (Patton, 1990). Challenges, however, have been made to both of these viewpoints. For example, in contrast to the centrality of theory, whether deductively or inductively produced, educational researchers such as Elbaz (1983), Schon (1983), Johnson (1987), Clandinin and Connelly (1996), Beattie (1997) and Elliott (1998) have made a case for the conceptualisation of teachers' personal practical knowledge. This is not applied theory but, in Johnson's terms, arises "out of our bodily experiences and provide[s] patterns that are meaningful to us and that influence our reasoning" (1987, p. 1). As Beattie explained, by understanding the teacher's knowledge as something dynamic and held in relation to practice, the teacher "whose knowledge is subjective rather than objective, fluid rather than fixed, embodied rather than cognitive, and relational rather than individual" (1997, p. 16) can be viewed as an autonomous agent in the educative process.

The varieties of research that have practice as a central focus are growing. Teaching as reflective practice (Schon, 1983), teachers as curriculum developers (Elliott, 1998; Stenhouse, 1975), action research (W. Carr & Kemmis, 1986), teacher research (Cochran-Smith & Lytle, 1990) and teachers' personal narratives (Clandinin & Connelly, 1996; Elbaz, 1983) all exemplify this trend. Teaching as reflective practice argues that the relationship between reflection and the development of professional knowledge is vital. Research on teachers' thinking acknowledges practice as thinking in action (Beattie, 1997), and notions of teacher as researcher on his or her own practice have increased in popularity. Stenhouse gave prominence to the idea of the teacher as researcher (McNiff, 1988).

His central message for teachers was that they should regard themselves as researchers, as the best judges of their own practice, and then the natural corollary would be an improvement of education. (McNiff, 1988, p. 25)

Stenhouse's idea was that of an educational science in which each classroom was a laboratory and each teacher a member of the scientific



community. Stenhouse, and later work by Kemmis, Elliott, McNiff and others, developed action research methodology. Underpinned by the work of social psychologist Kurt Lewin, action research engages people in inquiries into their own lives through a spiral of steps, each including four stages of planning, acting, observing and reflecting.

Another approach to research through teachers' personal narratives asks the question of how the embodied, narrative, relational knowledge teachers carry autobiographically and by virtue of their formal education, shapes, and is shaped by, their professional knowledge context (Clandinin & Connelly, 1995). However, as Shulman pointed out, despite a desire for teachers also to be researchers, both traditional process-product and interpretive/action research varieties of research have been carried out predominantly by university researchers rather than teachers themselves (Bruce Ferguson, 1999). Furthermore, others (for example, Bruce Ferguson, 1999; Cochran-Smith & Lytle, 1990) point out that teachers themselves, who have daily access, extensive expertise and a clear stake in improving classroom practice, have no formal way to make their knowledge of classroom teaching and learning part of the literature on teaching (Cochran-Smith & Lytle, 1990). Early in the research process I began to question my position as an outsider in what I had thought of as an action research project. While the teachers in both schools were immersed daily in assessment, and in a position to reflect and improve their practice, I could only facilitate reflection and document their changes.

In addition to these perspectives on researching teachers' practice, some educational researchers in the 1990s believed that just as there is no neutral education to be studied from the outside, there is no neutral research (Lather, 1991; Middleton, 1998). Lather explained that we are in a postpositivist period in the human sciences, which is marked by much methodological and epistemological ferment (Lather, 1986). This state of affairs has freed up space for the construction of new research designs. For example, in searching for research methods that allow us both to understand the maldistribution of power and resources underlying our society and to change that maldistribution in ways that

help create a more equal world, Lather has argued for research as praxis (Lather, 1991, p. 51). With roots in feminist research, neo-Marxist critical ethnography and Frierean participatory research, the notion of research as praxis brings together the use of theory to illuminate lived experience with the right of persons to participate in decisions that claim to generate knowledge about them. But, as Lather states, her goal is not to construct one new approach but “to move research in many different and, indeed, contradictory directions in the hope that more interesting and useful ways of knowing will emerge” (Lather, 1986, p. 272).

Having more interesting and useful ways of knowing would indicate that there are multiple perspectives in both research and practice. Eisner and Peshkin (1988) stated that what we also have is multiple perspectives on the state of methodology. In addressing the difficult question about how to assess the value of qualitative work, they believed that

Because there are no algorithms, no statistical significance tests for qualitative studies, that most exquisite of human capacities must come into play, judgement. But good judgement is not a mindless activity. It depends upon attention to detail, sensitivity to coherence, appreciation of innuendo, and the ability to read the subtext as well as the text. (Eisner & Peshkin, 1990, p. 12)

This stimulus to experiment, document and share efforts towards emancipatory research is stated by Polkinghorne thus

What is needed is for practitioners to experiment with new designs and to submit their attempts and results to other participants in the debate. The new historians of science have made it clear that methodological questions are decided in the practice of research by those committed to developing the best possible answers to their questions, not by armchair philosophers of research. (Polkinghorne, 1983, p. xi)

The research reported in this thesis was designed and carried out to find the best possible answers to questions initially raised by teacher practitioners in primary schools. In this research I also considered myself to be a practitioner, and thus set out to explore these questions with the help of teachers to enable the teachers to change their practice in constructive ways, to develop theory and illuminate it (Bassey, 1999), and to enhance my knowledge and practice as a

teacher educator in the sphere of educational assessment (Blacker, 1998). Hence, outcomes that changed practice during the research were anticipated, as well as an enhanced understanding of both the theory and practice of assessment in self-managing primary schools.

Furthermore, my interest in the topic of teachers' assessment practices, as I have already stated, sprang from a desire to assist teachers to resolve the tensions between accountability and informing learning and teaching. My understandings of the nature of competing educational and assessment discourses, explored in Chapters Two, Three and Four, as mechanisms of power (Blacker, 1998; Foucault, 1977; Middleton, 1998) which are diffused and not to be found at any one central site, urged me to concentrate on power's "microphysics".

A far more promising theoretical response is to concentrate on power's "microphysics". Since power is not homogeneous and can be approached only at its specific sites of application, one should not expect to gain a useful theoretical purchase at the macro-level.... Instead, the proper level of analysis is a worm's eye view; power must be considered from the "bottom up". (Blacker, 1998, p. 356)

Such a view requires a focus on local problems and local solutions. As I have explained in Chapter One, I have come to see universals as problematic and have found Foucault's notions of discourse, power/knowledge, surveillance and discipline helpful in understanding why teachers find assessment such a vexed area of their practice. The research questions set out in the conclusion to Chapter Four, together with the forgoing discussion of research methodologies, influenced me to proceed with a predominantly case study approach in this investigation.

### **A Participatory Case Study Approach**

Educational case study is a prime strategy for developing educational theory that illuminates educational policy and enhances educational practice (Bassey, 1999; Merriam, 1988). Cohen and Manion (1989) observed that "present antipathy towards the statistical-experimental paradigm has created something of a boom industry in case studies" (p. 125) using a diverse range of techniques

employed in the collection of both qualitative and quantitative data. In qualitative studies

investigators use a case study design in order to gain an in-depth understanding of the situation and its meaning for those involved. (Merriam, 1988, p. xii)

Rather than a methodology, case study research is a choice of what will be studied (Stake, 1994). In his reconstruction of educational case study, Bassey (1999) defined the process as one in which critical enquiry is aimed at informing educational judgements and decisions in order to improve educational action. Specifically, he explained that an educational case study is a study of a singularity conducted in depth in natural settings. Its essential features are that sufficient data are collected for researchers to be able to explore significant features of the case and to put forward interpretations for what is observed. These can be theory seeking/theory testing case studies, story-telling/picture-drawing case studies or evaluative case studies (Bassey, 1999). In a theory seeking/theory testing case study, the singularity is chosen because it is in some way typical of something more general. The focus is the issue as much as the case, and the enquiry is carried out in order to understand the issue of concern. Stake (1995) refers to research such as this, into one or more particular situations in order to try to understand an outside concern or issue, as “instrumental” case studies.

Bassey argues that “fuzzy propositions” (more tentative) and “fuzzy generalisations” (less tentative) can be developed in theory seeking and theory testing case studies as a way to communicate research findings to those who may use them. For him, story-telling and picture-drawing case studies use narrative accounts and “portrayal” to illuminate theory. Evaluative case studies, by contrast, set out to explore some educational programme, system, project or event in order to focus on its worthwhileness (Bassey, 1999).

Bassey warns that categorisation “is a dangerous game” (1999, p. 64), because these types of investigation seldom fit one category or another. In reality they overlap or have features outside of those described. With these reservations in mind, I used notions of theory seeking/theory testing and story-telling/picture-

drawing educational case study as a research strategy with which to address the research questions raised for investigation.

Theory seeking and theory testing case studies are particular studies of general issues (Bassey, 1999). The case is chosen because it is expected in some way to be typical of something more general (Bassey, 1999; Stake, 1995). By focusing on the complexities of educational practice through providing portraits to illustrate different approaches, comparative analysis can lead to important modifications of both practice and policy (Crossley & Vulliamy, 1984). Making a distinction between the kinds of case study is important because the methods used will be different depending on the purpose and type of case study (Bassey, 1999; Stake, 1995). “Methods”, here, refers not simply to data collection methods, but also to such matters as methods of analysis, theory generation, generalisability and emancipatory participation. In order to write about the methods of enquiry used and the way in which this study was conducted, I have divided the process into a series of research stages. In summary, the stages are

1. identifying the research issue and questions;
2. determining ethical guidelines;
3. collecting and storing data;
4. generating and testing propositions;
5. interpreting and illuminating the findings.

Although, in reality, these overlapped each other, they are presented here as if they were addressed in order for the purpose of simplifying the account for the reader. Each stage is discussed in detail below.

### Identifying the Research Issue and Questions

As I have explained earlier, my interest was in the ways in which managerial and educational discourses were impacting on primary school teachers in self-managing schools after the new national curriculum and assessment guidelines had been introduced. In particular, I wanted to know

- What effects had the repeated claims about failing schools had on teachers in classrooms?

- How did the new urgency and anxiety about standards, as a matter of government concern, intersect with professional educational discourses about assessment and pedagogy?
- To what extent did the proliferation of discourses about assessment result in a re-disciplining of teachers and students?
- To what extent had the curriculum and assessment requirements for monitoring students' learning disciplined teachers' educational beliefs and practices?
- How were these discourses taken up, used and resisted by teachers?
- How had these discourses of education and schooling impacted upon and disciplined my own practices as a teacher educator?

### Research Questions

Keeping these issues in mind, I identified three major research questions that guided the investigation. These were

1. What assessment discourses influence teachers' assessment practices in self-managing schools?
2. How did the teachers respond to these assessment discourses?
3. What effects did these discourses have on their teaching practice?

Because the research questions are "the engines that drive the train of enquiry" (Bassey, 1999, p. 67), it followed that I needed to pursue my investigation in schools which were addressing assessment issues. As I have indicated, this was not difficult as most schools at this time were wrestling with these issues. However, I did not initiate the research by approaching any schools because it was important that they were wrestling with assessment for their own purposes and that my involvement be at their request. Due to my previous work in this area (for example, Faire et al., 1994; Faire & Yates, 1994), two schools approached me in this regard early in 1996.

Following initial discussions with the principal of each school, both agreed to allow me to pursue my investigation with the teachers in their schools as I supported them in their endeavours to improve their assessment, recording and reporting practices. We saw this as a situation in which the school and teacher

development process could inform the research and, at the same time, the research might be informed by, and have an impact on, practice. Before entering the schools, however, I addressed the question of research ethics.

### Determining Ethical Guidelines

Bassey (1999) suggests three headings under which to discuss the ethics of research: respect for democracy; respect for truth; and respect for persons. Under respect for democracy, he argues that in a democratic society, and subject to the responsibilities imposed by the respect for truth and the respect for persons, researchers can expect certain freedoms.

The freedom to investigate and to ask questions; the freedom to give and to receive information; the freedom to express ideas and to criticise the ideas of others; and the freedom to publish research findings. (Bassey, 1999, p. 74)

Respect for truth, in which researchers are expected to be truthful in data collection and analysis and the reporting of their findings, has to do with ensuring the trustworthiness of these aspects.

### Trustworthiness

Trustworthiness is being able to show that the research is believable and is important in establishing the research's scholarly acceptance (Merriam, 1988; Morgan, 1981). Establishing trustworthiness involves ensuring, to the best of one's ability, that the findings, and the interpretations from the findings, actually reflect what was there (Morgan, 1981). Other terms for this process, such as "validation" and "verification" (Strauss & Corbin, 1990) and "validity" (Argyris, 1993), also refer to the notion of testing or judging the trustworthiness of the outcomes. In narrative research the term "authenticity" carries similar meanings and is essential to the credibility of the work (Beattie, 1997).

Trustworthiness also implies credibility of the information. Or, in other words, a way of validating the research (Lather, 1991). In this sense, Cronbach stated that the

job of validation is not to support interpretation, but to find out what might be wrong with it.... To call for value-free standards of validity is a contradiction in terms, a nostalgic longing for a world that never was. (1980, pp.103-105, cited in Lather, 1986, pp. 269-270)

Validation and finding out what might be wrong with an interpretation necessitates taking steps to continually interrogate the research process as well as the data. Bassey (1999) draws from the work on trustworthiness by Lincoln and Guba (1985), outlining a series of questions that might be used to ensure the trustworthiness of the information throughout the research process. He sets them out as follows.

Collection of raw data

1. Has there been prolonged engagement with data sources?
2. Has there been persistent observation of emerging issues?
3. Have raw data been adequately checked with their sources?

Analysis of raw data

4. Has there been sufficient triangulation of raw data leading to analytical statements?

Interpretation of analytical statements

5. Has the working hypothesis, or evaluation, or emerging story been systematically tested against the analytical statements?
6. Has a critical friend thoroughly tried to challenge the findings?

Reporting of the research

7. Is the account of the research sufficiently detailed to give the reader confidence in the findings?
8. Does the case record provide an adequate audit trail? (Bassey, 1999, p. 75)

As these questions suggest, continual checking throughout the investigation was necessary. For example, after all of the interviews, transcripts were returned for verification and used as a basis for initiating the subsequent interviews. Other data were also sought from outside the case study schools to



check that the data resonated with other teachers' experience. The backwards and forwards nature of the research in this thesis between researcher and teacher/s, the give and take, were employed in order to achieve trustworthy data that minimised the distorting effects of personal bias. In this way the tacit (subjective) and propositional (objective) knowledge might be interwoven and become mutually informing (Heron, 1981, cited in Lather, 1986). This reciprocity of approach was designed to provide the setting for a mutual negotiation of meaning and power in participatory research. Lather believes that reciprocity "operates at two primary points in emancipatory empirical research: the junctures between the researcher and the researched, and data and theory" (Lather, 1991, p. 56). Following Fay (1977, cited in Lather, 1986), Lather proposed several overt strategies necessary for full reciprocity in research. In summary, these were

- Interviews conducted in an interactive, dialogic manner, that require self-disclosure on the part of the researcher.
- Sequential interviews of both individuals and small groups to facilitate collaboration and a deeper probing of research issues.
- Negotiation of meaning. At a minimum this entails recycling description, emerging analysis, and conclusions to at least a sub-sample of the participants.
- Discussions of false consciousness...[where], in the nexus of that dialectic, lies the opportunity to create reciprocal, dialogic research designs which not only lead to self-reflection but also provide a forum in which to test the usefulness, the resonance of conceptual and theoretical formulations. (Lather, 1986, p. 266)

Reciprocity was built into this investigation in order to strengthen the trustworthiness of the data and theory building by constantly returning the findings to the participants for confirmation. In fact, the recursive nature of this study forms the basis for how it has been reported. Chapters Six to Nine move backwards and forwards between the school and teacher stories, explaining how interactive, dialogic interviews were conducted. All of the interview transcripts were returned to the teachers concerned for verification, confirmation, elaboration and comment. They were asked to respond to me about the transcripts, and when I did not receive them back, I followed up at our next meeting, to see if they had

read the transcript and whether it conveyed what they had wanted to communicate. Towards the end of the data-collecting process I had not heard back from two of the teachers to whom I had sent transcripts for checking and comment. Upon investigation I found that both had moved schools. They were still within the local area and I was able to confirm with them, verbally, that they were satisfied with the transcripts. One teacher edited her transcripts extensively, mainly to correct her grammar, but most edited only to clarify their meanings. Conference papers and journal articles written throughout the process of the investigation were sent to the schools and teachers concerned for critical comment and review before presentation. Feedback was sought on school-wide models from schools that had used them and such comments were taken on board.

Many times the teachers reported what Lather has coined “the click of recognition” (Lather, 1986, p. 271) when they were asked for comment on description and theory building, confirming credibility. For example, the term “head noting” was coined early in the research process to refer to the way most teachers held assessment information mentally about aspects of student achievement. It became an important aspect of theory building in this thesis and teachers almost always nodded vigorously as the term was unpacked. But in addition, this feeding back of initial interpretations and information analysis to the participants for their comments resulted in what Lather refers to as “catalytic validity” which

Represents the degree to which the research process reorients, refocuses, and energises participants towards knowing reality in order to transform it, a process Friere...terms conscientization. (Lather, 1986, p. 272)

Conscientization in this sense is the process of researcher and participants collaborating in theory building and critique for the purpose of transforming educational situations towards the improvement of education (W. Carr & Kemmis, 1986; Lather, 1986). Rather than researcher neutrality, this approach to producing theory with the help of the participants can overcome, to some extent, the lack of relevance much research has for teachers. Sometimes this occurred at my request, such as when I wanted to discuss data from the previous interview and see if I had

interpreted responses as they were meant. But more often this was at the request of the teachers. I met with the assessment review team from Waimaka School early on in the project to debate their renewed understanding of how a school-wide system for assessment should work, and together we thrashed out an approach for them to try. On another occasion one of the three teachers whose practice was studied in depth rang me and asked if I would go and talk about how his assessment practice had changed as a result of moving into a junior school classroom.

Trustworthiness in this study was not about trying to obtain some detached sense of objectivity or validity, and very much about making it clear where the researcher stood and how my point of view was modified and substantiated. For example, working with the three teachers who became the focus of the investigation became very time consuming as we met to discuss our findings about their practice. One teacher asked if we could meet outside the school situation so we could compare and contrast individual styles of teacher appraisal that she felt had impacted on various teachers' assessment and recording practices. She did not feel comfortable critiquing this with me within the school setting, but we came to understand the impact of the syndicate (a group of teachers at one class level who work together to plan programmes and appraise each other) and appraisal processes on practice. Another teacher began (and continues) to contact me to come and see something new in action and hear about his changed practice.

Some commentators are not enamoured with the term "trustworthiness" and do not accept it as a valid criterion for guiding or judging qualitative research (Wolcott, 1994). For example, Scheurich contends that trustworthiness is just one of the masks worn by validity

that conceal an underlying sameness, a singularity of purpose or function that transgresses the supposedly incommensurable differences or boundaries dividing various research epistemologies. (Scheurich, 1997, p. 80)

Scheurich's discussion of validity, within his attention to research in the postmodern, tackles the issue of how, as a researcher, he (or any researcher) can engage with the otherness of the other without transforming her (or him) into

purely one's own. He argues that "developing new imaginaries of validity...is difficult" (Scheurich, 1997, p. 88) and, after reviewing and critiquing alternatives such as collaboration, attentive care (White, 1991) and Lather's "transgressive validity", he contemplates silence as the best alternative.

From my perspective, in carrying out an educational case study which aimed at informing educational judgements and decisions in order to improve educational action, trustworthiness appeared to be the most practical approach and congruent with my beliefs about the research process. In practice, Wolcott (1994) suggests that the researcher who is concerned about trustworthiness should be a good listener who records meticulously all aspects of the research process and seeks critical feedback from others (in this case, especially the teachers). Furthermore, the researcher needs to be very open about how he/she goes about the research, hiding nothing – including weaknesses and limitations (Strachan, 1997; Wolcott, 1994).

### Dependability and Consistency

In the same way that trustworthiness adds to the credibility of participative research, readers should be able to recognise the categories and findings in other, similar situations. This does not mean, as in quantitative studies, that the results should be replicable (Merriam, 1988), but that the processes of data gathering and analysis are systematic enough and robust enough to stand up to scrutiny and be replicable. Moreover, it needs to be clear that rather than presenting a general picture, the findings of a study such as this are context specific and consistent with that context. Some argue that a better term for this is "coherence" (Beattie, 1997).

In order to establish the dependability and consistency of this study several techniques were used. Firstly, information about teachers' assessment practices was gathered from many sources (Lincoln & Guba, 1985; Merriam, 1988). For example, although initially a total of 12 teachers from the two participating schools were interviewed, further interviews took place with 20 teachers from primary schools throughout New Zealand. These schools were selected for their distribution throughout the country, across the decile rankings (ten socio-economic

levels by which schools are categorised in New Zealand) and due to their urban and rural locations (see Chapter Seven). In this way it was possible to check out whether the information gathered from the initial two schools was idiosyncratic, pertaining maybe just to those two schools or schools in the Waikato, or whether it was indicative of some wider trends and was of consequence to other teachers.

As well as gathering information from different sources, a range of methods were used to collect the information. Using multiple methods of data collection is usually referred to as “triangulation” (Merriam, 1988). As outlined below, I used interviews and observations of practice, field notes were collected as I participated in the assessment development process, and I was able to use the school records compiled by the teachers as sources of information.

Thirdly, the research process was described in detail (see Chapters Six to Eight) so that others, both researchers and teachers, could follow the process and see how and why decisions were made and categories arrived at (Wolcott, 1994). Merriam (1988) and Bassey (1999) refer to this as providing an “audit trail”.

Finally, it was important to describe clearly the context of the research, the theories that underpin it and my own relationship with the research participants (Merriam, 1988). Merriam refers to this technique as the “investigator’s position” (1988, p. 172), but others (for example, Battersby, 1981; Harold, 1995) use the term “researcher’s predispositions” to refer to the strategy of explaining the assumptions and theories behind the study, the researcher’s position vis-à-vis the group being studied, the basis for selecting the participants and the social context from which the data are collected (Merriam, 1988).

### Predispositions of the Researcher

Some argue that in undertaking qualitative research the more the researcher knows about a situation, the more difficult it can be to study it as a researcher (Spradley, 1979). In fact, Ball (1985) cautions that involvement can be such that the researcher actually becomes a participant. Battersby (1981) also warns that bias on the part of the researcher is a danger and suggests that “an honest assessment by the researcher of his or her dispositions and how these influence the

methods used and, in turn, the findings, can mitigate its effects to some extent” (p. 186). As explained earlier, though, rather than an objective, neutral stance, my role as a facilitator was central to the reciprocal nature of the research. I did become a participant in the process at times but the trick was to be fully aware that this was so and to gather data from participants with possible differing viewpoints.

I had not taught in either of the case study schools although I had been involved as a consultant in one when they had carried out an earlier review of their assessment procedures. In addition, two teachers in one of these schools had been participants in a course of mine that was designed to assist the participants to improve their assessment knowledge and practices. One of these teachers had also attended one of my courses about junior school teaching. To an extent this must have affected the research context. These two teachers knew me and something of my beliefs and attitudes about assessment, learning and teaching. My influence, however, in the context of multiple and competing discourses of assessment argued in this thesis, was but one voice and, in terms of courses run, had had the potential to affect teachers in many schools throughout the region.

### Gaining Ethical Approval

In a desire to achieve respect for persons, I recognised the teachers’ initial ownership of the data and endeavoured to respect them as fellow human beings who are entitled to dignity and privacy. I took steps to ensure this was carried out in practice. Initially, approval was sought from and given by the research ethics committee at the School of Education. An essential requirement for carrying out human research in our institution, this process served to heighten my awareness of the ethical behaviour required. This process, in combination with my own integrity, professionalism and personal accountability (Merriam, 1988), highlighted the following critical ethical considerations.

Firstly, no harm should be done to the participants because of their involvement in the research process (Bogdan & Biklen, 1982; Spradley, 1979). To address this concern, I informed all of the possible participants about the nature of the research and requested their informed consent in writing before involving them

in the research (see Appendix A). As well, the reciprocal nature of the research (Lather, 1986) helped to protect the research participants from exploitation as they were involved in checking and confirming the results throughout the research process. The teachers were assured in writing that they could withdraw from the project at any time and were never under pressure to continue to take part. This ethical guideline was implemented when one teacher declined to be interviewed during the study due to pressure of workload, and investigations at one school were suspended during a particularly stressful period so that potential harmful effects might be avoided.

Maintaining confidentiality and anonymity are also essential ethical concerns in human research. There were two groups within this research to whom this consideration applied: the participating teachers and the children within their classes. Confidentiality was maintained in this study by assigning code names to the schools and the teachers involved. In fact, in the two case study schools the teachers chose their own code names in most instances. This made it possible for me to return transcripts to the appropriate teachers for checking while protecting the confidentiality of the information at all stages of the project. The use of code names also protected the privacy of the children by not identifying their teachers and by using pseudonyms for the children in the occasional case where their teacher referred to them. I discussed whether it would be necessary to obtain informed consent from the children and/or their parents with the principal of each school and board of trustees. They advised me that, because the focus of the research was sharply upon the teachers and their practice, and because pseudonyms would be used for the schools and the teachers, the privacy of any children was not an issue and further informed consent would not be necessary. As well, the research data and tapes were stored confidentially.

Thirdly, because the research process was intended to be reciprocal and, to some extent, emancipatory for the teachers, an open research agenda was desirable (Lather, 1986). This meant that the participants were kept informed about the progress of the research, conference papers and articles were circulated to the participants concerned before publication and the research report was made

available to those who took part (Spradley, 1979). Because research reports are often written for an academic audience, I also made a point of publishing material for teachers in teacher journals as a result of the school development aspect of the study (M. Hill, 1997b; 2000), spoke at teacher conferences (M. Hill, 1997a; 1998b; 1999) and gave oral feedback throughout the investigation (Strachan, 1997).

An open research agenda also allowed for change in direction when this appeared necessary, either to protect the participants involved or to better pursue the purpose of the research. This took place on several occasions during the study. For example, some of the original participants moved schools during the research process. In some cases it was appropriate to follow them to their new locations to continue with the research, while in others it was not. Decisions were made on a case-by-case basis, keeping in mind possible disruptions and the importance of each contribution to the research itself.

Difficulties can arise when the researcher's interpretations are at odds with those of others (for example, those of the participants or others connected with the study). The temptation for the researcher may be not to give the same weight to the participants' views that she or he gives to her or his own (Peshkin, 1988). The recursive structure of this study, returning repeatedly to participants for critical comment on both the data and the interpretation, was intended to address this concern. While there are multiple interpretations, the researcher's own being one of these, the aim is to use the recursive structure of the research design to come to an agreed position about the findings, one that will be useful in understanding better the situation of the teachers and in illuminating their practice.

Finally, there was the possibility of some ethical and copyright issues in this research because of my involvement in undertaking contract action research and school development contracts for the Ministry of Education (M. Hill, 1998a) concurrently with the investigation. This meant that I needed to be very clear about when I was gathering information for my thesis and when I was not. I sought and received permission from the Ministry of Education to carry out interviews with teachers at some of the same schools involved in the action



research project in recording and reporting (M. Hill, 1998a). These interviews are explained in Chapter Seven. I also ensured that any materials I used in the school development programme that I was directing in 1998-99 that had their source in this (thesis) research were exempt from the copyright clause in the contract with the Ministry of Education. Furthermore, I was careful to obtain informed consent from all my research participants, including those who became involved through their participation in my contract work for the Ministry of Education. (See Appendix A for copies of informed consent formats and letters.)

### Collecting and Storing Data

As stated earlier, initially I set out to undertake action research case studies to achieve change-enhancing enquiry (Bassey, 1999; Bruce Ferguson, 1999). W. Carr and Kemmis define action research as

A form of self-reflective enquiry undertaken by participants in social situations in order to improve the rationality and justice of their own practices, their understanding of these practices, and the situations in which the practices are carried out (1986, p. 162).

But I believed that there were contradictions in using an approach that put teachers in the central research role when a major purpose of my research was both theoretical and descriptive. As Henry and McTaggart explain

Classroom action research typically involves the use of qualitative, interpretive modes of enquiry and data collection by teachers (often with help from academics) with a view to teachers making decisions about how to improve their own practices. Primacy is given to teachers' self-understandings and judgements. The emphasis is "practical", that is, on the interpretations teachers and students are making in the situation. If professional researchers are involved, their role is a service role to the teachers. They are often advocates for "teachers' knowledge" and may disavow the relevance of more theoretical discourses such as critical theory. (Henry and McTaggart, undated, p. 6, cited in Bruce Ferguson, 1999)

It seemed important that I be able to act within these schools as both a facilitator and an academic researcher. W. Carr and Kemmis suggest three main levels of engagement in action research for outsiders, such as myself: technical, practical and emancipatory. Technical action research is used to describe studies in which the facilitator co-opts “practitioners into working on externally-formulated questions which are not (necessarily) based on the practical concerns of teachers” (W. Carr & Kemmis, 1986, p. 202).

Practical action research, in contrast, involves outside facilitators forming co-operative relationships with practitioners and helping them to participate in the research cycle outlined above. This second type is so named because it assists participants to develop their practical reasoning. The facilitator’s role in the practical type of action research is Socratic. That is, she or he provides a sounding board against which practitioners can try out ideas and learn more about the reasons for their own action and about the process of self-reflection.

At the emancipatory level of action research, rather than the facilitator assuming the Socratic role, the participants themselves take on the responsibility for self-reflection and analysis. The facilitator in this type of research can be described as that of a moderator who can help form a self-critical and self-reflective community, but, once established, allows the community itself to sustain and develop its work (W. Carr & Kemmis, 1986). This notion of participative emancipatory research implies a sense in which teachers are seeking more control over their practice, and searching for “ways of knowing which interrupt relations of dominance and subordination” (Lather, 1991, p. xvii). This emancipatory level of action research seemed appropriate to the work the teachers had asked me to participate in with them.

I worked in partnership with the teachers in the two schools that had asked me to assist them as they reviewed and changed their assessment policies, systems and practices. In return, some of the teachers also agreed to participate in my thesis project, allowing me to record their development processes, conduct in-depth interviews and observe their assessment practice in action. Although still an “outsider” in these schools, I became purposefully involved in the development

process and was able to gain access with the full agreement of the staff but in an unobtrusive manner (Spradley, 1979).

It was while I was working alongside the teachers in these two schools in the emancipatory action research manner that I discovered that my purpose and emerging methodology were beyond the scope of this approach alone. As Strachan (1997) warned, “the process of qualitative research is not always neat and straight forward” (p. 87), and researchers often need to be flexible enough to cope with the inevitable design and redesign process. Because qualitative methods are flexible and evolving, in order to gather relevant information (Patton, 1990; Spradley, 1979), I extended the research design beyond emancipatory action research to include semi-structured interviews, documentary analysis and extended observations of classroom practice and the analysis of school records. Rather than using an action research methodology, this investigation became a participatory case study in which data were gathered using qualitative methods and analysed inductively. The methods used were predominantly interviews, observations and the analysis of official texts and school records.

### Interviews

Lincoln and Guba (1985) described interviews used in enquiry as ranging from structured to unstructured. In structured interviews

the problem is defined by the interviewer before the interview...  
the questions are formulated ahead of time and the respondent is  
expected to answer in terms of the interviewer’s framework and  
definition of the problem: (Lincoln & Guba, 1985, p. 155)

In contrast, unstructured interviews have a non-standardised format and the interviewer is concerned to learn as much as possible about the individual viewpoint of the respondent. An unstructured interview stresses the interviewee’s definition of the problem or situation, encourages the interviewee to structure the account, and relies on the interviewee’s notions about what is relevant to the situation and what is not (Lincoln & Guba, 1985). In making a similar point,

Kvale (1996) used two contrasting metaphors to illustrate different kinds of research interviews: interviewer as miner and interviewer as traveller.

In the miner metaphor, knowledge is understood as buried metal and the interviewer is a miner who unearths a valuable metal. Some miners seek objective facts to be quantified, others seek nuggets of essential meaning.... The alternative traveller metaphor understands the interviewer as a traveller on a journey that leads to a tale to be told upon returning home. The interviewer-traveller wanders through the landscape and enters into conversations with the people encountered.... The traveller may also deliberately seek specific sites or topics.... What the travelling reporter hears and sees is described qualitatively and is reconstructed as stories. (Kvale, 1996, p. 3-4)

While the “miner” interview is often of the structured variety mentioned above, the interviews for the “traveller” research might best be termed semi-structured (Fontana & Frey, 1994; Kvale, 1996). Kvale (1996) explains that a semi-structured interview has a sequence of themes to be covered as well as suggested questions. He adds that, “at the same time, there is an openness to changes in sequence and forms of questions in order to follow up the answers given and the stories told by subjects” (p. 124). Within this type of interview, as well as leading others to new understanding and insight as they, through their own storytelling, come to reflect on previously natural-seeming matters of course in their own culture,

The journey may not only lead to new knowledge: the traveller might change as well. The journey might instigate a process of reflection that leads the interviewer to new ways of self-understanding, (Kvale, 1996, p. 4)

I saw myself very much in the traveller role of interviewing. Although the interview questions for each of the interviews were prepared in advance, there was a great deal of scope for the interview to follow the respondent’s agenda, with room to pursue an unexpected or interesting aspect as it arose (see Appendices B, C and D). The questions in the first interview (see Appendix B), with 12 teachers in the two case study schools, were very general in order to encourage teachers to share as much of their knowledge, attitudes and practices of assessment as

possible. The subsequent interviews with these teachers, however, were more focused in order to learn more about particular aspects, check out conclusions developed from earlier data or understand particular ideas in more depth to work towards reciprocity (Lather, 1991; Lincoln & Guba, 1985). A third round of interviews was undertaken with seven of the initial 12 teachers. In addition, as described in Chapter Seven, 20 further interviews were carried out with teachers in other schools throughout New Zealand.

In order to encourage reciprocity, the interviews were conducted in an interactive, dialogic fashion that both invited response and required self-disclosure on the part of the interviewer (Lather, 1991). This allowed me to negotiate meaning and recycle previous description, analysis and conclusions back to the teachers for comment and re-examination. This style of interviewing also paid dividends in terms of establishing close rapport with the teachers, essential to “opening doors to more informed research” (Fontana & Frey, 1994, p. 367). In the two case study schools where the teachers were sequentially interviewed over a two- to three-year period, a sense of trust was gradually built up, especially with the three teachers I observed in action. For example, one of these teachers approached me later to act as both a referee and appraiser. Another continues to keep in touch, seeking advice about studying for further qualifications. Middleton (1988) reported interviewing each woman in her research up to four times in this way in order to follow up and clarify issues raised in previous interviews, and, in the process, developing close and lasting friendships. Oakley (1981) refers to the relationship built up through reciprocity of this nature as one of intimacy.

All of the interviews were audiotape recorded. Although some of the teachers felt a little self-conscious about this initially, most appeared to forget that this was happening soon after they began. Tape recording the interviews left me free from note taking so that I could give my full attention to the discussion. This enabled me to follow the thread of the conversation, often covering the intended questions in different order, probing for more information or following an interesting lead. As well, it allowed me to be a good listener, respond sensitively,

and attend to other signals such as body language and non-verbal cues (Kvale, 1996; Lincoln & Guba, 1985).

However, as Lincoln and Guba (1985) warn is often the case, there were times when mechanical problems interfered with taping, or, as Kvale (1996) reminds, the interview situation is so novel or complex that the researcher forgets to turn the tape recorder on or the background noise makes the tape inaudible! Inevitably such calamities occurred. Batteries ran out, the dishwasher in one staffroom drowned out most of one interview and, on another occasion, the tape recorder simply failed to tape the interview at all.

Most of the interviews, though, were successfully recorded. I developed a useful taping kit that included a small, high-quality cassette recorder that was voice activated and had a visible light that indicated when it was recording. This was invaluable as it meant I could check inconspicuously whether it was working without interrupting the flow of the conversation to rewind and replay just to make sure. I invested in a separate condenser microphone that boosted the volume and, I'm sure, improved the sound reproduction tremendously. The kit also contained spare batteries, extra blank tapes, labels and a spare pen.

Although I began by transcribing two of the early interviews myself in order to check the acoustic quality and review my own interviewing style, the majority of the tapes were transcribed for me by several experienced dictaphone typists. Kvale suggests that although transcribing "seems like an apparently simple and reasonable procedure, transcriptions involve a series of methodological and theoretical problems" (1996, p. 163). For example, some (such as, Charmaz, 1995) advise that the researcher herself or himself should transcribe all the interviews. This facilitates familiarity with the contents and alleviates problems such as how to make sense of some statements. However, I decided not to do all the transcribing myself. I was inexperienced with dictaphone typing and a slow typist. The fact that I was using a qualitative analysis computer program (NUD\*IST), which required insistent re-reading of all the transcripts, was an added factor in having the tapes transcribed for me. The transcriptions were not perfect. They contained mistakes, sometimes because the typist could not

understand what was being said, sometimes just typing errors. I re-listened to the relevant tapes and manually corrected the transcripts. After I had corrected any obvious errors in the transcripts, I sent them back to the teachers to check and comment on, and to expand or delete material they didn't think appropriate. Some teachers heavily edited their transcripts, reworking them into written language, while others made no alterations at all. Frequently the teachers commented on changes they had made since the interview took place, or as a result of it.

Another more theoretical issue is that once transcriptions are made, they tend to be treated as *the* solid empirical data in the interview project (Kvale, 1996). As he puts it

Transcribing involves translating from an oral language, with its own set of rules, to a written language with another set of rules. Transcripts are not copies or representations of some original reality, they are interpretive constructions that are useful tools for given purposes. (p. 165)

Kvale suggested that the question "What is a correct transcription?" cannot be answered – as there is no true, objective transformation from the oral to the written mode. He suggested that a more constructive question is "What is a useful transcription for my purposes?" (1996, p. 166). In my study I was concerned to make sense of the teachers' practices, beliefs and ideas about assessment and, therefore, although the transcripts were used for analysis they were seen as unfinished explanations and, to some extent, partial conversations. I often began the next interview by referring back to the last and asking for explanations and further analysis of a dilemma or tension that had arisen from my reading of an earlier transcript. I became aware of the nuances and the differences, the contextual meanings, and made notes to assist me to keep these in mind when retelling the stories.

Lincoln and Guba (1985) recommend that as the study proceeds the information accumulated from interviews should be subjected to critique in later interviews. This was done in the study reported here, both with the same teachers and with a national sample of teachers in a range of types of schools. In addition to three interviews with each of the three case study teachers, each was also

observed for a week in action in her or his classroom and the information in the interviews was crosschecked with the observation material. This was one way in which the information collected was triangulated to help in establishing its trustworthiness (Kvale, 1996; Lather, 1986; Patton, 1990).

### Observations

The observational continuum extends from, at one extreme, the situation where the observer is completely uninvolved in the action, possibly monitoring it at timed intervals, to, at the other extreme, the involvement of the researcher as a fully participating member of the action (Adler & Adler, 1994). Between these poles of the continuum there are many possible stances the observer can take. I decided to use “peripheral membership” (Adler & Adler, 1994), which involves becoming part of the group being studied, to the extent that an insider’s perspective is important to understanding the situation. Although there is close interaction with the group’s members, the researcher is not involved in the core activities of the group.

A peripheral stance was chosen for the classroom observations of the teachers because it seemed the most appropriate way to help answer the research questions set out above. In detail, the observation involved shadowing each teacher for a full week of teaching. Shadowing involved being with each teacher all day and following him or her around as she or he worked, both inside and outside the classroom. In order to make an accurate record of the daily assessment activity the teacher wore a tape recorder and the tapes were constantly recording the interactions she or he had during the day. As well, I was usually close enough to hear the interactions and took running records, noting each activity as it occurred, with details of what was happening, and marking with asterisks and notes in a separate column of the field notebook any possible evidence of assessment-related activity. All the tapes were numbered and noted in the field notebook so that crosschecking between the field notes and tapes was possible.

There can be drawbacks in this method of data gathering. For example, Adler and Adler (1994) warn against faking and observer effect. When people are



observed (and interviewed) they can change their behaviour, doing and saying things that they normally would not (Ramsay, 1987). Using observations of classroom practice to check out interview data (Strachan, 1997) and continuing the observations over a reasonably lengthy period of time (Wolcott, 1973) can both help to minimise these effects.

This style of observation allowed me to become part of the action and, within a short time, the novelty of my appearance in the classroom seemed to wear off. Although the teachers all felt self-conscious initially, all reported that they soon forgot the tape recorder was there most of the time, and unintended interactions were often taped. This had benefits because private communications (with, for example, other teachers or teachers' family members) could be erased later but other events, not initially recognised as assessment activities, could be retrieved for analysis from the tapes. I tried to steer clear of "going native" (Cohen & Manion, 1989, p. 129) (becoming involved in the core activity of the classrooms) and mostly avoided requests for help (in order to concentrate on what the teacher was doing), but, at times, this was impossible. For example, in a new-entrant classroom (5-year-olds new to school), several children emptied all the sand from the sandpit one very hectic Friday afternoon. As the teacher was involved in assisting children to clean up the classroom, I took over managing the refilling of the sandpit, much to the teacher's relief!

Another bonus of using the tape recorder during the observation period was that it could be used to capture action when I was not on the scene. This occurred one morning when I was unable to reach the school before the teacher started work. Of her own accord, she obligingly turned the tape recorder on before she began the day's early planning and administrative tasks, and talked them into the record as she went, indicating also that she was at ease with the research process.

The observation tapes were not all transcribed, as it was not necessary to use all the taped information. The running record and field notes assisted me to find portions of the tape that were of interest and use, and these I transcribed and used as text in the same way as I have described above for the interviews.

### School Records

In a study of teachers' assessment practices it seemed to be important to examine and include information retained by the teachers as records of their assessment practice. As Gipps et al. (1995) note, teachers' assessment knowledge is often tacit and difficult to access through interviews, but all teachers are required to keep assessment records in some form. And, as Hakim (1993) points out, administrative records, such as school records, are used to record the development and implementation of decisions and activities that are central to their functions. However, consistent with my use of interviews and observations in this study, these assessment documents were not treated as isolated pieces of evidence and used for content analysis as in more positivist models (Jupp & Norris, 1993), but rather were used as artefacts of the assessment practice which could be treated as text and examined for their role in the research process. In Foucault's terms, I began to ask "What do these records *do*?" as far as learning and teaching, reviewing and reporting go (Blacker, 1998).

As I interviewed the teachers and observed their practice, I asked permission to copy samples of the records they created and kept. These ranged from notes on scraps of paper, through anecdotal comments on children's work and highly organised checklists, to books created by the teacher specifically for tracking learning and the teacher's own programme. I stored these alongside the interview transcripts and observation notes and transcripts as aids to interpreting practice.

### Generating and Testing the Analytical Statements

Initially, in order to begin to explore the data I was gathering about the assessment practices and beliefs of the teachers in the two case study schools, I began analysing the tapes as soon as I had undertaken the interviews. I also made an immediate start on analysing the field notes I had taken during staff meetings. I developed codes and categories of codes to assist me in this process, in much the same way described by Charmaz (1995) as grounded theory (Strauss & Corbin,

1990). From these findings I generated propositions about what seemed to be happening and returned to interview the teachers again to ask further questions.

Because the interview transcripts became the “raw files” of data for the computer analysis programme, they needed to be typed in a particular format. The typists used a large font (usually Times 14) and set the transcript out with returns after each speaker finished. Code names were used for each of the participants to ensure confidentiality during the transcription process.

Due to the large number of interview transcripts I used QSR NUD\*IST software to assist me with the analysis. NUD\*IST stands for Non-numerical Unstructured Data Indexing Searching and Theorising. This software programme was specifically designed to deal with qualitative data in a way that allows a researcher to deal with large amounts of descriptive data in a reasonably systematic fashion without becoming overwhelmed (Kvale, 1996; Middleton & May, 1997). In this project I was dealing with 50 interviews and 33 tapes of classroom observation recordings, in addition to many, many pages of field notes and memos. In the past, I had resorted to using highlighters, cutting and pasting transcripts, and using index cards to sort information into categories (M. Hill, 1998a). I had also used computer software to categorise data but, because I had been analysing survey data, I had used pre-set categories following from the questions to do this. My search for the ways in which teachers responded to discourses of assessment in their classrooms across the curriculum meant that I needed to find a way of building the analysis from the data.

Using a qualitative software package, however, does not mean that the transcripts are fed into a black box and emerge magically transformed into manageable theory (Rouse & Dick, 1994). On the contrary, using NUD\*IST entailed a very systematic approach to introducing each transcript (or document) into the programme and then coding segments of the documents, usually each response to a question. The benefit of using this program was that, once coded, NUD\*IST could retrieve all or any of the like-coded information, make reports on each node (category), and allow the coded data to be arranged hierarchically into chunks of analysis. NUD\*IST manages data documents, allowing easy retrieval

and editing, and creates an indexing system that can be investigated independently of the data. Above all, NUD\*IST is designed for asking questions and building and testing theories (Qualitative Solutions and Research Pty Ltd, 1997).

NUD\*IST is well suited to, and frequently employed for, grounded theory (Rouse & Dick, 1994; Selvaraj, 1996; Strauss & Corbin, 1990). Whether used in positivistic or interpretive ways, the characteristics of grounded research which unite the empirical research process with theoretical development often underpin qualitative research processes. Essentially, these characteristics comprise

- simultaneous involvement in the data collection and analysis phases of the research;
- the development of analytic codes and categories from the data;
- the development of middle range theories to explain behaviour and processes;
- memo-making or analytic notes as an intermediate step between data coding and writing first drafts of papers;
- theoretical sampling; that is, sampling for theory construction, not for representativeness of a given population, to check and refine emerging conceptual categories;
- and the delay of the literature review. (Charmaz, 1995)

I used these strategies in combination with the NUD\*IST program. I was still interviewing teachers as I started to read, transcribe and code the first taped interviews. New questions, prompted by the data I had collected, emerged during this process, and I returned to interview all the 12 teachers a second time to clarify earlier information and to check out the propositions I was generating as I went about my analysis of the data. I kept a research journal with memos continuously throughout the empirical process. These ranged from draft propositions to initial theorisation, from ideas for further questions to proposed methods to explore a new lead. My decision to interview teachers nationally to check out the propositions I had developed from the initial interviews and observations fits with the notion of theoretical sampling (Charmaz, 1995; Strauss & Corbin, 1990). And,

while I did start early on my literature review, I returned to it with new ideas later in the research process.

However, in contrast to grounded theory building, where the relationship between data and theory is that theory follows from data rather than preceding it (Charmaz, 1995; Glaser & Strauss, 1967; Lather, 1986), I began this investigation with theoretical perspectives. As elaborated in Chapters Two, Three and Four, I understood the educational assessment context in which the teachers were operating as contested. I argued that multiple discourses appeared to be competing for their attention, and, as Weedon (1987) has pointed out, some of these appeared to have greater weight and power than others. Thus, although I incorporated strategies characteristic of grounded theory, my approach was not typically that of a grounded theory study. There were at least two major points of departure.

Firstly, as indicated in Chapter Four, my reading of the literature on assessment research had alerted me to work in England and Canada that had illuminated understanding of teachers' classroom assessment practices and identified broad categories of practice, some more productive of learning than others. I interrogated my data for signs of similar approaches. Likewise, theories of formative assessment in New Zealand classrooms, developed by Bell and Cowie (1997), provided me with ways to analyse in some depth the formative strategies described and observed in the classrooms I investigated. This was more than theoretical sensitivity (Strauss & Corbin, 1990). I actively set out to connect these theoretical perspectives with my own.

Secondly, some researchers, as interested as I am in the everyday relations of power in institutional settings (such as teacher-student exchanges), have seen the potential of Foucault's work for analyses of spoken discourse (Comber, 1997; McNeil, 1986; Meadmore, 1993; 1998; Middleton & May, 1997; Walkerdine, 1984). Comber (1997) found, in relating everyday talk to macro-discursive practices, that "the metaphor 'traverse of discourse'...(was) a helpful device in considering how power is exercised in Foucault's 'capillary' fashion in workplaces such as schools" (p. 6).

Similar statements, constructions, images and metaphors may appear and reappear in written and spoken texts like the policy

document, the syllabus, the principal's memos, the staffroom conversation, the teachers' guidebook, and, of course, the textbook. (Luke, 1995, p. 371, cited in Comber, 1997)

By considering both the discursive practices of those who make authoritative statements about the role of assessment in the New Zealand curriculum and the discursive practices of the teachers as they appeared in both spoken and written texts, I endeavoured to identify recurring vocabulary that might signal assessment discourses. Published official texts, such as curriculum documents, *Education Gazette* statements and ERO reports, were read with and against the everyday texts produced in the course of teachers' work. In treating the transcripts as texts, and reading them with and against the official texts, I explored the contradictory nature of teachers' work and the multiplicity of assessment discourses produced in these schools at that time. Words such as "accountability", "levels", "monitoring", "records", "requirements", "barriers to learning", "achievement", "standards", "checklist", "child-centred" and "progress" appeared repeatedly across the assemblage of national policy documents, interviews, classroom interactions, official school documents and teachers' assessment documentation. These words and terms seemed to be used and interpreted differently by different teachers, signalling intersections between managerial and educational discourses. I began to focus on them as key words. Because I had read that NUD\*IST can be used for handling data for all types of qualitative research (Richards & Richards, 1994), including qualitative research conducted within a theoretical framework – for example, Miles and Huberman's qualitative data analysis (Rouse & Dick, 1994), or life-history narratives (Middleton & May, 1997) – I explored its use with key words.

### Interpreting and Illuminating the Findings

Having explored my data, generated and checked propositions, and theoretically sampled to check for resonance with a wider group of schools and teachers in New Zealand, I believed it was necessary to illuminate my interpretations by investigating a few teachers' practice in far more detail.

Following other investigations into the connections between teachers' personal narratives and their professional practices (for example, Clandinin, 1998; Comber, 1997; Munro, 1998; Pryor & Torrance, 1996), I studied practice in order to develop a theory of practice. Rather than viewing practice in terms of theory, using narratives about teachers' practices shifted the interpretive process from only my perspective to a more mutual researcher-participant reconstruction of meaning in action (Beattie, 1997). Following Pryor and Torrance (1996), I undertook classroom observations in order to witness these deeply embedded learning and teaching encounters and discuss them with the teachers involved, thereby enhancing the trustworthiness of the findings.

### Narrative reporting

Bassey explains that "narrative writing is usually best organised in sections in order to help the reader" (1999, p. 87). Chapters Six to Eight in this thesis each represent a period of time during which a significant stage of the research was conducted. These sections are in chronological order, reporting how questions were asked and answers sought, and how propositions were tested and modified and explanations sought. Data collection, analysis and interpretation are intermingled rather than reported as results.

Within these sections are contained both teachers' stories of assessment, recording and reporting and school stories about school-wide assessment, recording and reporting systems. Clandinin and Connelly (1995) use the notion of stories as narratives of experience, "both personal – reflecting a person's life history – and social – reflecting the milieu, the contexts in which teachers live" (Clandinin, 1998, p. 2). As well, they used the term "personal practical knowledge" to capture the idea of experience in a way that allows talk about teachers as knowledgeable and knowing persons. Furthermore

Keeping our eyes firmly on the question of teacher knowledge we realised that knowledge was both formed and expressed in context. Within schools this context is immensely complex and we adopted a metaphor of a professional knowledge landscape to help us capture this complexity. (Clandinin, 1998, p. 3)

These terms and metaphors seemed particularly appropriate as an aid to understanding how the discourses of assessment impacted on teachers in these New Zealand primary school classrooms. As in Clandinin and Connelly's work, the teachers in my work talked of two fundamentally different kinds of places in school: one inside the classroom, with (and without) their students; and another outside of the classroom, with other teachers, the principal and others such as the ERO and national policy makers. I have, therefore, adopted these terms and metaphors within this thesis as an aid to analysis and theory building as well as a way to tell the story(ies).

In Chapter Eight three portraits are presented in a manner described as "portrayal" (Bassey, 1999). Following Stenhouse (Bassey, 1999), portrayal is an attempt to preserve some of the qualities of narrative in descriptive writing that lacks a storyline.

As in documentary film, characters, incidents and descriptions of an environment in which they are set are juxtaposed to provide a portrayal which is interpretive of the case as a whole. (Stenhouse, 1985, p. 52, cited in, Bassey, 1999, p. 88)

Gipps et al. (1995) used vignettes (small portraits of teaching) as a research tool in investigating teachers' assessment practices. While I didn't use these as a specific research method, I did use portraits to give the research participants and, eventually, the interested reader a sense of the complexities of practice, an understanding of the setting and a sense of being there. The teachers participating in the research very quickly identified the portrait most similar to their own style of practice and offered information to make it more realistic from their perspective. They corrected my interpretation when necessary and added their reasons for their own practice as they went.

### Timing

The fieldwork phase of this investigation took place from 1996 until the end of 1998. Therefore, it began following the implementation of self-managing



schools and the *New Zealand Curriculum Framework*, including the 1993 revised National Education Guidelines (see Chapter Three, pp. 68-69), and was completed prior to the proposed trial of externally-referenced testing of whole cohorts of children (Ministry of Education, 1999a) and the publication of the revised NAGs (Ministry of Education, 1999b). Since its completion, the proposed trial of externally-referenced testing has been abandoned. Knowing about the timing of the investigation is important because, as I have argued in Chapters Two, Three and Four, ruptures and breaks in the systems of reasoning (in this case national assessment policies and regulations) tend to cause changes in local sites of practice (teachers' assessment practices in classrooms). For example, new National Education Guidelines, published after the fieldwork for this thesis was completed, but not yet made compulsory at the time of writing, have already begun to affect practice, as have new national initiatives to supply all schools with information technology capable of recording children's progress.

### **Summary and Comment**

In summary, the research design used for investigating the research questions outlined above employed a participative, case study approach that was responsive to the need to gain access to appropriate research contexts. As explained above, trustworthiness, dependability/consistency and ethical considerations were acknowledged and addressed as important elements of this type of research. I believed that a form of case study design in which, to some extent, I was a participant, was the most suitable research genre for this investigation because of the rich, thick, descriptive nature of the data that were necessary to explore the teachers' knowledge and actions that were at the heart of this study. Furthermore, I considered the study's recursive nature essential for examining in depth the complexities of assessment in action in schools and classrooms.

Because I was investigating how teachers in self-managing schools responded to the competing discourses of assessment and the effects these

discourses had on their teaching and assessment practice, several suitable methods of enquiry became apparent from the research questions. For example, to investigate what the assessment discourses were that influenced the teachers' assessment practices in self-managing schools, it seemed necessary to ask the teachers and senior staff in each school through interviews, to analyse their written records and to listen to their discussions about assessment. Observing the teachers in action was a useful strategy to use to investigate how the teachers responded to these competing assessment discourses and the effect the discourses had on their teaching practice. Therefore it appeared necessary to use multiple methods to achieve a clearer picture of the action.

This investigation mainly took place in two Waikato schools. This was the context in which the in-depth interviews, school development and classroom observations took place: in other words, these were the schools I “hung about” in. But I also checked out whether the information gathered from these two schools was idiosyncratic, pertaining maybe just to those schools or schools in the Waikato, or whether it was indicative of some wider trends. There was movement backwards and forwards between Waikato and national contexts throughout the duration of the investigation as the stories of assessment emerged. For reasons of clarity, I have presented the research in these contexts in three phases. In Chapter Six, phase one – setting the scene – introduces the assessment stories about the two case study schools. The second phase, which explored competing discourses and their effects on teachers' assessment practices, is reported in Chapter Seven. The third phase of this investigation, an examination of the assessment practices of three teachers in their classroom contexts, is reported in Chapter Eight.

## **CHAPTER SIX**

### **SETTING THE SCENE: ASSESSMENT STORIES FROM TWO SCHOOLS**

Getting started in case study research involves deciding on the entity to be studied. But one cannot define what the case will be without first having a research problem. (Merriam, 1988, p. 41)

The journey may not only lead to new knowledge; the traveller may change as well. (Kvale, 1996, p. 4)

The “problem” (Merriam, 1988, p. 41) – or as I think of it, the matter of doubt and uncertainty – that prompted this investigation was teachers’ concerns about assessing, recording and reporting student learning. More specifically, I wanted to know how teachers in self-managing primary schools were responding to the competing discourses of assessment and what effects these discourses were having on their teaching practices. As I have explained, my methodology sent me on a journey of discovery that focused on teachers at work in schools. This chapter sets the scene for the unfolding stories of assessment that I constructed about these teachers from 1996.

The stories begin in two Waikato schools and move backwards and forwards between school stories and teacher stories through my interactions with the teachers in interviews, through school-based action research and through classroom observations. These stories did not, however, unfold neatly. Far from it. The process seemed more one of unravelling, searching, digging and scaffolding. At times there seemed to be more pieces missing than found, and I have had to become a narrator who, once I thought I had the story right, returned time and again to try it out on those I had first heard it from.

The first section of this chapter introduces the two Waikato schools that participated in this project and reveals how I began to hear their stories of assessment. In the second part of the chapter I reflect on what I learnt about

assessment from these school and teacher stories and identify multiple and, at times, competing discourses of assessment. The chapter concludes by briefly reviewing the early findings and raising further questions.

## **Stories from the Schools**

### **Arawhata School**

The first contact I had with the participants in this study was a telephone call from Terry, the principal of Arawhata School.<sup>5</sup> He phoned me early in 1996 and asked whether it would be possible for me to work as a consultant in his school to assist the teachers to clarify and improve their assessment practices. He had phoned me because I had worked with this school several years before and he knew I had been working with other schools on a professional development contract in Assessment for Better Learning (ABeL) (Calder, 1996). At that time I had been working on my PhD proposal and had planned to investigate how schools and teachers were responding to assessment requirements in the Tomorrow's Schools environment. I explained that I was trying to cut back on my consultancy work with schools so I could begin my research project but that I would consider his request and get back to him.

At first I was agitated. I had been attempting to make progress with my doctoral work for several years and kept putting it off to work with schools and teachers, policy makers and education advisers on assessment issues. I was determined to make progress but also felt morally impelled to continue working with teachers as they were obviously wrestling with issues of assessment, standards, accountability and documentation. I discussed the request with a senior colleague and together we decided that maybe this was just what I had been looking for. I rang Terry back and asked to come and talk to him about working in the school with a joint purpose: on the one hand as a facilitator in the school's

---

<sup>5</sup> Arawhata is a pseudonym used to protect the school's anonymity.

change process; and, on the other, as a researcher. He agreed and we met to discuss how we might go about it.

At the meeting I talked with the senior management team, consisting of the principal (Terry), deputy principal (Brian) and assistant principal (Julie), as well as Alex (my senior colleague), about what the school was wanting and how my research might fit alongside their plans. Terry, Brian and Julie were concerned that, although they themselves were clear about the assessment expectations in the school, many of the teachers were anxious and seemed confused about what the school required in terms of assessment. The senior management team believed that, as outsiders, Alex and I might provide a fresh perspective on the assessment scene at the school. To get things under way, Alex and I agreed to interview a range of staff members about their knowledge and understanding of assessment policies and practices at the school and report back to the senior management team on our findings.

At the meeting we also talked about my plans for researching assessment policies and practices as a result of the restructuring which followed the Tomorrow's Schools reforms. We convened a second meeting to talk with the rest of the staff about both the senior staff's intentions to carry out a school review of assessment, recording and reporting, and my request for them to take part in my research. Following this staff meeting I sent a letter detailing my request to undertake the research and enclosing an information sheet and consent forms for the principal, board of trustees and teachers involved to sign (see Appendix A). Much to my delight and relief, all agreed to take part in the research.

Arawhata School was a large Years 1-6 primary school with students aged between 5 and 11 years. The school roll numbered approximately 575 students and there were 26 full-time teachers on the staff at the time. It was situated in a suburb of a large provincial city and was categorised as a decile 9 school.<sup>6</sup> Eighty

---

<sup>6</sup> Schools are ranked in one of ten decile groups to indicate their need for extra funding based on the socio-economic status of the surrounding community. Decile 1 schools receive the most funding, decile 10 schools the least.

three per cent of the children were classified as European, 11 per cent as Maori, four per cent as Asian and two per cent as Pacific Island.

The teachers who agreed to become participants in my research were:

**Terry:** the principal. He initiated the contact but took leave from his job for most of 1996 to work with an education consultancy company.

**Brian:** the deputy principal. He did not have a class. Brian was acting principal for most of 1996.

**Julie:** the assistant principal. She did not have a class. Julie took leave from her job to work in another area of education from August – December 1996, and after returning at the beginning of the 1997 year, left teaching in August 1997.

**Lisa:** a first-year teacher with a Year 3 class. Lisa had successfully completed an advanced course in classroom assessment and evaluation co-taught by me the previous year.

**Sam:** in her seventh year of teaching with a Year 2 class.

**Anne:** a very experienced teacher with a Year 1 class.

**Rose:** a mature woman with a young family who had recently returned to teaching and was teaching a Year 2/3 class.

**Rosemary:** a second-year teacher with a Year 4 and 5 class.

**Chris:** an experienced male teacher, recently appointed as a senior teacher with a Year 5 class.

The ERO had carried out the last school effectiveness review in 1994. The report stated that the board of trustees should consider

- utilising the achievement information staff collated for this review as a starting point for further development;
- aggregating and analysing student achievement information collected on entry to school to establish a baseline from which to measure progress;
- ensuring that assessment information was recorded for all students in all curriculum areas.

The school's action plan for 1996 included the objective "to review assessment procedures". Under "actions", it stated, "to employ a person with the

knowledge of education assessment and needs of teachers and children, to review our school class assessment procedures". As Terry explained it, the senior management team wanted our help to investigate how the teachers were really feeling about the assessment system at the school and what improvements might be necessary. Alex and I set about interviewing a sample of 14 teachers at the school (see Appendix B). Some of the teachers were from the management team while others varied in experience and the class levels at which they taught.

In general we found that there was a clear expectation that they needed to assess all areas of the curriculum, but that how and how much were up to the individual teacher. There were a few exceptions to this. The teachers reported that two running records (of reading) were expected per term, that information be recorded in the students' individual profiles twice a year, and that many of the teachers in the senior school (Years 4-8) were confused about whether it was mandatory to carry out Progressive Achievement Tests<sup>7</sup> (PATs) or not. Two of the teachers felt that they didn't get enough guidance with respect to assessment.

Without a doubt the teachers felt that observations of the children were the most valuable assessment activities. They used these to find out where they were at so that they could make decisions about where they had to move to next. Many also saw this as "good" assessment, assessment for a clear purpose. The teachers stated that they liked to keep their own informal notes about children's progress. One teacher valued pre- and post-tests and running records most because they gave hard data.

In response to a question about the assessment activity they considered to be of least value, most stated it was the profiles. Typical responses were: "It's done for the next teacher and I know she probably won't use the information – even if she bothers to read them."; "It's very daunting..."; "They're over the top." The kindest interpretation was that "...they are done for Julie and she wouldn't ask for them if she didn't need them". One teacher believed that tracking spelling

---

<sup>7</sup> Standardised norm-referenced tests for Years 3-10 students in listening comprehension and for Years 4 – 10 students in reading comprehension, vocabulary, mathematics and study skills. At the time of this research, some of these tests were undergoing revisions. See also Chapter Two.

progress was a waste of time, while another felt that although goal setting was “trendy”, it was of little value.

In fact goal setting seemed to be quite a contentious issue. It was not clear whether goal setting meant students setting their own goals or teachers goal setting with or for students. In response to a question about their views on student goal setting, self- and peer assessment most teachers separated these aspects out in their answers. While two teachers were strongly in favour of student goal setting, another was tentative and another very strongly opposed. Widely different practices were also noted, with some undertaking (time-consuming) daily goal setting with children, some setting weekly goals and others goal setting once a term with the child and his or her parents. Some focused on social and behavioural goals while others focused on academic goals or a combination of both. It was interesting that many believed student self-assessment was worthwhile but didn't use it. Peer assessment was not commented upon.

None of the teachers we interviewed had seen the school's policy on assessment or remembered having any input into it. One thought the profile folder was the policy and another remembered seeing a folder of policies. Although many of the teachers we interviewed were new to the school, others had been there several years. It seemed that the assessment expectations were passed on verbally between teachers or prompted by requests for specific information (for example, the profiles or school reports) by management staff at certain times of the year.

In response to being asked for suggestions for professional development in the assessment area, many suggested that it would be a good idea to make assessment easier and more “child-friendly”; they believed that it would be good to avoid duplication, inject some new ideas and introduce student self-assessment.

We reported these results to the senior management team in writing and met a second time with the senior management team to discuss the results and talk about how they might go about professional development. At the meeting, the senior management team decided to hold a development day for all the teaching staff on a Saturday in late June. The purpose of the day would be threefold. Firstly, to provide a forum for open discussion about the current policies and



practices in assessment. The senior management team agreed not to attend this part of the day to allow the rest of the teachers to speak as freely as they wished about the current situation. Alex and I agreed to lead this session and designed an activity that encouraged the teachers to record their ideas about the current assessment procedures to facilitate this aspect of the day.

In the second session, Brian, the deputy principal, read the staff a picture book *Seven Blind Mice* (Young, 1992), which exemplified the idea that although each teacher knew about assessment, there was not a clear unified picture of what was expected in the school and that this would be the objective of the review they were beginning that day. Alex shared the results of our interviews with the staff and led them through a session that clarified assessment terminology and examined the Ministry of Education requirements (the NEG's and the NAG's). In the final part of the day, Brian and Julie invited the staff to examine what was working and what needed improvement in assessment at Arawhata School.

At this time there was common agreement that teachers were expected to fill in (over each student's time at the school) a 21-page booklet which, in effect, was a profile of development in each curriculum area. Several running records were expected to be carried out for each child in reading each year. A system of sampling was used throughout the school in order to monitor progress across selected areas of the curriculum. To achieve this, groups of ten children in each year group were selected. Each group was representative of different abilities and gender/ethnicity. Tasks were set, carried out and marked, and the deputy principal and assistant principal analysed and reported on the results school-wide.

In the junior school (up to and including Year 3) every child was assessed at school entry (within 4-6 weeks of their fifth birthday) with a school-designed entry check. Data were kept and analysed from this information. There was a consistent policy of undertaking the Six Year Diagnostic Survey (Clay, 1979) with children as they turned six. If they were reading at or above a particular level, a full running record check was undertaken. If they were below the level set, a full six-year-old diagnostic survey was carried out. In Year 3 the PAT in listening was

used with about 60 per cent of the students. Teachers used their judgement about who might take the test.

In the senior part of the school (Years 4-6), all children sat the listening PAT, running records were used twice a year to determine reading ages,<sup>8</sup> all sat the mathematics PAT, and it appeared that most used pre- and post-tests for units in mathematics. There were targeted programmes for children with special needs and special abilities, organised and overseen by the deputy and assistant principals. Written reports were provided for parents on every child's progress and achievement twice yearly and there was a system of reporting achievement to the board of trustees.

However, there was no clear statement of exactly what was expected of teachers by the school in terms of assessment and its documentation. The result of the development day was a wish list for the assessment review. It included

- the modification of profiles to a manageable size;
- the desire for little or no duplication of records;
- the need for the assessment policy to be a working document;
- a request for guidelines for reporting to parents;
- a request for a set of guidelines for school assessment procedures; and
- a clear idea of who each document was for and a desire that the procedures meet the Ministry of Education requirements for assessment.

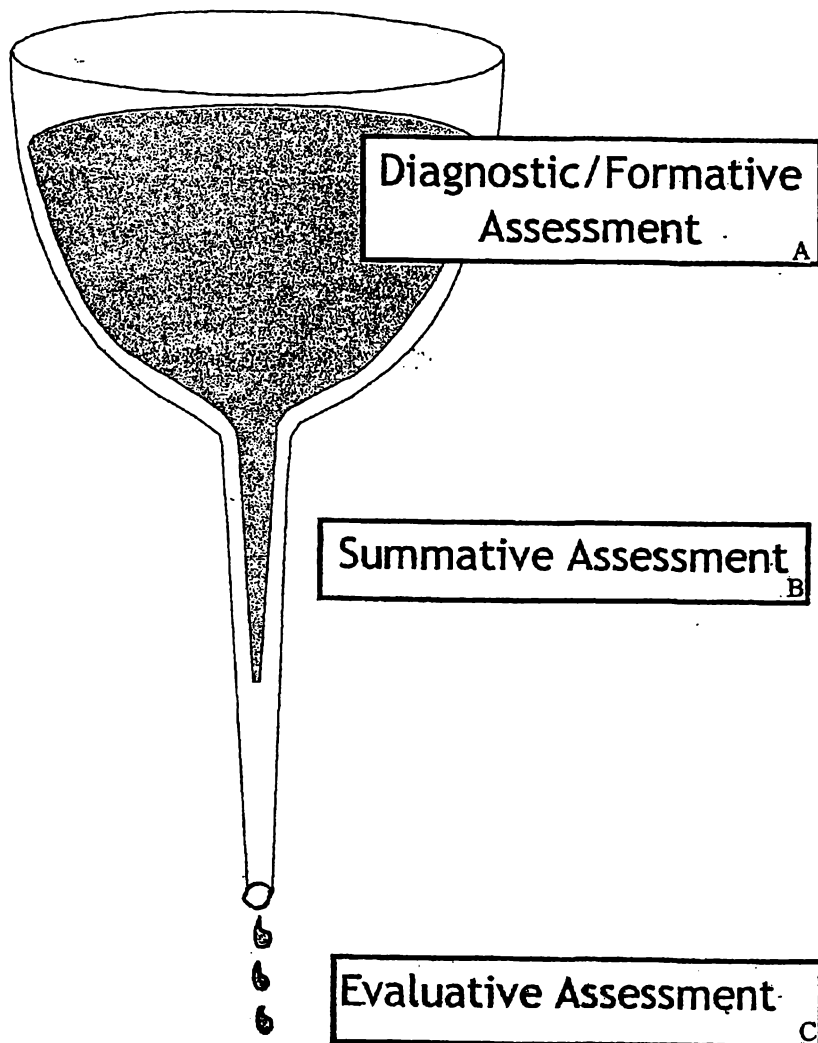
Brian also asked if any of the teachers would like to join the senior management working party to form an assessment group to action the outcomes of the day's work. Several teachers expressed a desire to be involved and the day closed with an open invitation for teachers to participate in any of the future meetings.

Over the next month this assessment working party met to clarify assessment at Arawhata School. They invited me back early in April to respond to a proposal they hoped to present to the whole staff. We squeezed into the principal's office and gathered around the whiteboard on the wall. On it was a

---

<sup>8</sup> It is worth noting here that determining reading ages is not a stated purpose of the reading running record procedure but is a widely held and practised misinterpretation of it.

large diagram of a funnel (see Figure 6.1 below). The group explained the funnel to me as a metaphor for their recording and reporting practices. They had used a funnel because they decided that it indicated the relative importance and effort they felt should be attached to each aspect of the system.



**Figure 6.1** Funnel Model of School-wide Assessment generated by the teachers at Arawhata School

The conical part of the funnel (A) represented the data that each teacher collected and used about learning during day to day teaching. The group

suggested that a roll book<sup>9</sup> be supplied to each teacher for this aspect of assessment and recording, and that the teachers be free to use this as it suited them except that it should contain information about children's progress across all the curriculum. As they explained it, the roll book would be a place to collect "field notes" (Brian) on children's learning. The intention was that it be manageable and functional for each individual teacher but that it should be retained in the school should the teacher leave as one important source of evidence about children's learning. It was envisaged by these teachers as the "context for day to day data" and as available as a source of information for parents if necessary.

The lower part of the funnel (B) represented the summaries teachers wrote from time to time about children's progress. The teachers saw this as necessary, but the smaller portion of the funnel allocated to it indicated the lesser importance they believed summaries should assume. They visualised these summaries as contained in a profile folder fulfilling two functions: the first was as a place where past and present work samples and assessment task results might be housed; the second was a double-sided tracking sheet which should be filled in periodically to indicate progress in learning.

The "drips" (C) issuing from the bottom of the funnel were envisaged as school-wide assessment information about the progress of cohort groups of children. The teachers believed that particular essential learning areas should be targeted for focus each year in their implementation plans and that the sample of approximately ten per cent of children at each year level assessed for progress and achievement in these areas should continue to be monitored. As explained above, this was already a well-established practice at the school and its continuance was supported at the staff development day. Although the drips from the funnel indicated that this was a small but important function of assessment at Arawhata School, they didn't show that the intention was that this school-wide information should be fed back into future programme planning as well as used to report to the board of trustees on the performance of the school. It was not clear whether this

---

<sup>9</sup> Roll books were similar to a register with children's names down the left hand column and columns ruled down the pages with headings for main topics at the head of each column.

was because it was not the teachers intention to do so, or whether it was just that the metaphor of the funnel did not lend itself to a recycling of this information. This was a major issue from my perspective and led, later in the investigation, to inquiries about how the assessment information was used.

Although there were still questions about what the tracking sheet for each individual student might track and how the data might be adequately recorded, the group decided to present the idea to a whole staff meeting the following week. They also agreed to divide the curriculum areas between the group's members and come up with a suggested format for each for the tracking sheet. Their criteria for the tracking sheet were that they wanted something simple and precise that allowed ongoing recording and included attitudinal as well as achievement information.

Within an hour at the whole staff meeting the funnel model was adopted. The staff made many constructive comments and agreed to the proposal that the ideas be turned into a draft handbook on school-wide assessment, that the tracking sheet be prepared, that roll books be purchased for the following year and that there be some sessions for staff to share ways these might be used during the rest of 1996. They agreed to continue with their present systems until 1997, including the school-wide monitoring that was already under way.

Suffice to say here that the stories the teachers told in response to our questions set me wondering. In contrast to the management team's view that the system needed fine tuning, several of the teachers seemed stressed and frustrated by the (lack of) school requirements. I wasn't sure that I was asking the right questions. Many seemed to confuse assessment and recording. Some were very stressed at just the mention of the "A" word. So later in April I decided to go back and ask more questions, try different angles and see if I could get deeper into the meanings that these teachers made of their assessment practice. I prepared a second set of questions (see Appendix B) and used these to interview in depth the teachers who had agreed to take part in my research.

At the time of this interview I was aware that the teachers were responding to a range of demands, some from the school, some from the ERO, others from the

professional development they were undertaking in the new curriculum documents and still others from collegial expectations. These demands were not necessarily complementary and, to varying degrees, often seemed in competition with each other. However, at this stage I didn't recognise these as multiple discourses of assessment. I could hear voices, echoes from my primary teaching days, complaining about standardised assessment demands, arguing against norming children as 1s or 5s, and struggling to fit school-wide tests into an already busy classroom schedule. But in addition I was hearing passionate stories about the effects of assessment on teachers' personal lives and conflicting messages about the worth of documenting evidence of every step in learning. These initial stories presented me with a rich source of information from which I needed to make some sense. As explained in Chapter Five, rather than extract what I took to be the meanings of these findings, I returned again to discuss these issues with the teachers in later interviews in order to facilitate co-construction. These issues are picked up again later in this chapter.

### Waimaka School

Buoyed by the beginning of my investigation at Arawhata School, I decided to wait a while before approaching another school to take part in case any others approached me for help. As explained earlier, my thinking was that if schools approached me it would be because they were committed to school review in assessment, rather than it being my agenda first and foremost. Late in April 1996 my patience was rewarded. Katy, the principal of a small rural school, rang to say she had recently been appointed and would appreciate my assistance with reviewing and improving their school-wide assessment systems. I had worked with Katy in her capacity as deputy principal at another school that had participated in the ABEL contract (Calder, 1996) the year before. When Katy was appointed to Waimaka School the board of trustees had made it clear that improving the assessment systems was a priority for their school. Their recent ERO report had recommended significant improvements and they knew Katy had

had experience implementing portfolio systems at her previous school and as an assessor for the National Education Monitoring Project (Flockton, 1999).

In retrospect, the fact that both of these schools had approached me to work with them to assist them to improve their assessment systems reflects the pre-existing relationship I had with them. To an extent this relationship meant that there was already a feeling of trust, at least between the principals and myself. We also had some shared understandings about learning, teaching and assessment. The same could be said for my relationship with two of the teachers from Arawhata School and one at Waimaka, whom I had known through professional development and university course work. However, I had not met the other staff members at Waimaka School before.

During my first visit to the school I discovered that Katy's request was motivated as much by the latest ERO report (that criticised the lack of a cohesive assessment and monitoring system) as by the teachers' desire for change. Katy and the other teachers explained to me at a staff meeting that they needed a manageable system that met the ERO requirements to "monitor student progress against the achievement objectives at all levels" and "assess student achievement, maintain individual records and report on student progress in all essential learning areas and at all levels" (Education Review Office, 1995b, p. 10). These requirements were set out in the school's 1995 assurance audit<sup>10</sup> as "actions required for compliance" and were credited in the audit as being National Administration Guidelines 1 (ii) and (v). However, they differed from NAGs 1 (ii) and (v) in that the words "in all essential areas and at all levels" had been added by the ERO.

---

<sup>10</sup> At the time this research took place, the ERO was undertaking two different types of school visit. One was the Assurance Audit. The purpose was to inform the Minister of Education, boards of trustees and the public about how well the legal requirements of the school were being met (Education Review Office, no date-a). The other, the Effectiveness Review, reported on the effectiveness of schools in relation to student achievement (Education Review Office, no date-c). Both types of report were open to the public. The ERO stated that they "are constructive and assist individual institutions to improve their performance. They are designed to promote accountability and high-quality education at the local and national levels" (Education Review Office, no date-b).

Waimaka School was located in a small rural settlement in the Waikato. As well as children from farming families, a significant number of children were from families employed by the local dairy factory. When this study began in 1996 there were 76 children at the school ranging in age from 5 to 13 years (Year 1 to Year 8). 72 per cent were of European origin, 26 per cent were Maori and 2 per cent were classified by the ERO audit report as “other”. This decile 4 school had received three visits from the ERO, the last being for a discretionary audit<sup>11</sup> in early 1996.

The teachers at Waimaka all agreed to become participants in my research. They were

**Katy:** recently appointed principal with 16 years’ teaching experience who taught a class of Year 6, 7 and 8 children.

**Marvin:** a teacher of at least 20 years’ experience teaching the Year 1 class.

**Rosie:** in her 3rd year of teaching. Rosie taught the Year 2 and 3 class.

**Luke:** also in his 3rd year of teaching. He held a long-term relieving position at the school and taught a class of Year 4 and 5 children. Luke moved to Kōtiritiri<sup>12</sup> School in 1997.

**Viv:** a part-time teacher with 10 years’ experience who released the principal during maths time each day and was also the reading recovery teacher.

I met with the five teachers of Waimaka School in the middle of May and talked with them about their goals for reviewing and improving assessment, recording and reporting. We also discussed my research plans and I sent a letter, information sheets and consent forms for them and the board of trustees to sign following the meeting (see Appendix A).

Soon after our initial meeting, I returned to the school and interviewed each teacher individually about assessment, recording and reporting. The

---

<sup>11</sup> Discretionary audits are visits carried out by the ERO to examine the extent to which a board of trustees has met its legal obligations and undertakings to the Crown with particular reference to areas of non-compliance identified in an earlier assurance report.

<sup>12</sup> Kōtiritiri School was a two-teacher school that became involved in this project in 1998 when I observed Luke teaching there for a full week. See Chapter Eight for further details.



responses to the questions in my first interview (see Appendix B) were strikingly similar to those gathered at Arawhata School. The valuable information gained from assessment was not necessarily that which was recorded. Many felt stressed even discussing assessment, never had enough time to devote to it and were unsure of exactly what they were expected to collect. They were dubious of keeping checklists and profiles of achievement, using standardised tests and documenting evidence of progress on lists of objectives. They felt that checklists and profiles were too time consuming and often not used by anyone and that standardised tests were similarly administered but not used. On a more positive note, these teachers also had strong opinions about the kind of assessment they found valuable: assessment planned for and carried out within teaching, as well as various tests and diagnostic tasks. Several referred to this type of assessment as “child-centred”.

All of the teachers, including the principal, stated that they would like to have the school requirements for recording and reporting assessment information clarified. Several also seemed to be duplicating their assessment and recording to meet school and the ERO requirements and their teaching needs. For example, one experienced teacher told me

Probably the assessment I believe is most useful and what I actually do doesn't marry up. I build [assessment] tasks into my units, that's something I feel is very useful, but I still can't get away from checklisting, which when I think about it afterwards, I wonder what is the purpose of doing it. (Marvin, Waimaka)

This situation had led to several of the teachers questioning who assessment was for. This questioning was framed in general terms, such as “Not just in this school, but you hear we've got to keep checklists for this and this and this, and it seems to be mind boggling and you wonder who it's all for” (Viv, Waimaka). All of the teachers at Waimaka School were keen to spell out for themselves how they should monitor progress and achievement at their school.

I then met with these teachers at a staff meeting, where they told me they had all agreed to keep a portfolio of samples of each child's work across the curriculum. This was, in some cases, a scrapbook with the samples pasted in, or in others, a clear-view folder with the samples filed in the plastic sleeves. The main

purposes of this portfolio appeared to be for periodically showing the child's progress to parents and as a memo device for the teachers when writing records and reports about individual students.

In addition to this portfolio the teachers wanted to develop a common way of keeping cumulative progress records about individual children. At the meeting they voiced different ideas about how this might be done. One teacher thought a computer software package would be useful, others were still adding information to the "pink cards"<sup>13</sup> (Department of Education, 1989c) and thought that they should design a replacement card to use throughout the school. All, however, had rejected the lengthy profiles that had been produced by the previous principal for this purpose as they were time consuming to fill in and seemed not to meet anyone's needs.

By the end of the meeting the staff had agreed to share examples of cumulative record card formats from various sources at their next meeting and then come up with a record card to meet their own purposes. They agreed that they needed something that was effective in tracking progress across the curriculum but not onerous to fill in.

When I next visited this school Katy showed me a draft copy of the record card they had designed. It had three double-sided pages. We talked about how it might be produced as the school did not have the equipment or expertise to produce a professional looking card. Katy decided to approach the School Advisory Service to see if a school management adviser could help with this process.

In addition to the information from the interviews, the school review processes I have described enabled me to learn a great deal about how teachers at these two schools were responding to the assessment challenges posed by both their constitution as self-managing schools and the national requirements to assess, monitor, record and report student achievement against the national achievement objectives. From the teachers' responses to my questions it became clear that in

---

<sup>13</sup> "Pink cards" refers to the Keeping School Records cards described earlier in Chapter Two.

attempting to meet these challenges they were drawing on multiple discourses of assessment.

## **Discourses of Assessment**

As I have explained earlier, discourses of assessment refer not to language or social interaction in general so much as to a relatively well-bounded area of pedagogical knowledge. According to Foucault's rethinking of discourse, in any given historical period we can write, speak or think about a given social object or practice (educational assessment, for example) only in certain specific ways and not others (McHoul & Grace, 1998).

A discourse would then be whatever constrains – but also enables – writing, speaking and thinking within such specific historical limits. And we deliberately speak of a “discourse” in the singular: for even though Foucault very frequently uses the mass noun “discourse”, he is typically keen to point out that this is something of a theoretician's shorthand, a way of signalling some common and general properties of discourses. Historically specific discourses are quite distinct from one another as well as from earlier and later forms of themselves which may or may not have the same names...they are discontinuous. (McHoul & Grace, 1998, 1998, p. 31)

Within the responses to my interview questions and the interactions that took place during each school's review of their assessment systems, multiple and competing discourses of assessment were evident. One group of assessment discourses appeared to relate to the ways in which teachers thought they should be accountable to various audiences, while another group related to how assessment could be child-centred and inform learning. For example, it appeared that teachers in both schools said they used assessment mainly as a tracking device for students' achievement. Although they used terms such as “child-centred” and said they needed to know where children were “at”, assessment as a means of changing their own teaching practice was not really visible in the responses to the interview questions. I have therefore addressed these as accountability discourses and formative discourses respectively.

### Accountability Discourses

As outlined above, both Arawhata and Waimaka Schools had initiated contact with me because they had determined that they needed to improve their assessment, recording and reporting systems. Although these schools were in some ways very different from each other, both principals believed that reviewing and improving the systems that kept track of children's learning and achievement was an important goal for their school. Teachers also seemed to see "good" assessment in terms of accounting for learning. For example, one teacher told me

Good assessment [is] when you do a unit, when you're setting three very broad objectives, at the end when you're seeing how well the children have actually achieved them, how well they can do it. So good assessment is nutting into exactly where the children are at the end of the unit and how far they've come since you began the unit and how much they've actually taken on board. (Rosemary, Arawhata)

And another explained that

good assessment should [provide] a level that when a child is leaving I can automatically fill out the pink card from using my roll book. It should be that [the principal] or someone else could use my roll book and get an appropriate picture of that person, where they're at in my classroom. (Luke, Waimaka)

From the interviews it appeared that the teachers also wanted to use the review process to clarify school-wide expectations, develop consistency among the staff in the way they monitored progress, make the systems useful, and, in one teacher's words, "make it user-friendly" (Rosie), rather than find ways to inform their own practice. Missing from these interviews was any sense that the teachers wanted to refine assessment to help them change their teaching and there was no mention of culture or gender in respect of assessment.

### Accountability to Governmental Requirements

In reviewing and improving their school-wide systems both of these schools were responding to the national requirements for assessment (the NEG's and the NAG's) and to the ERO reports they had received. As explained in Chapter Three, the NEG's and the NAG's set out the requirements for boards of trustees, through the principal and staff, to foster students' achievement through the delivery of the New Zealand Curriculum (Ministry of Education, 1994). While *The New Zealand Curriculum Framework* states that "the primary purpose of school-based assessment is to improve learning and the quality of learning programmes" (Ministry of Education, 1993a, p. 24), as I have argued in Chapter Three, NAG 1 seems to suggest that this can be done through a set of managerial practices.

At Arawhata and Waimaka Schools the teachers interpreted these regulations to mean that the progress of each individual student needed to be assessed and recorded as achievement (or otherwise) of the objectives set out in each of the National Curriculum Statements (in some cases this meant using the pre-1993 syllabus statements, as the new curriculum was still being produced). This interpretation was reinforced by the ERO reports each had received. At Arawhata School the 1994 Effectiveness Review Report stated that the board should consider

- utilising the achievement information staff collated for this review as a starting point for further development;
- aggregating and analysing student achievement information collected on entry to school to establish a baseline from which to measure progress;
- ensuring that assessment information is recorded for all students in all curriculum areas. (Education Review Office, 1994)

At Waimaka School the last Assurance Audit had listed 14 "Actions Required for Compliance" (Education Review Office, 1995b). Under the heading of curriculum management these stated

- implement learning programmes based upon the underlying principles, stated essential learning areas and skills *at all levels* (NAG 1 (i)) (emphasis added)
- monitor student progress against the national achievement objectives *at all levels* (NAG 1 (ii)) (emphasis added)
- assess student achievement, maintain individual records and report on student progress *in all essential learning areas and at all levels* (NAG 1 (v)) (emphasis added). (Education Review Office, 1995b)

It is interesting to note the inclusion in this report of the words “at all levels” and “in all essential learning areas and at all levels”, as these words are not included in clauses (ii) and (v) of the NAGs quoted above (Ministry of Education, 1993b, p. 4).

The implication of these recommendations and compliance requests is that by assessing student achievement, monitoring progress, maintaining individual records and reporting on student progress, learning and the quality of learning programmes will be improved. My observations and the responses to the initial interviews revealed that, on the contrary, the teachers mainly saw these recommendations as record keeping activity to be done over and above the assessment they regarded as most valuable for informing their teaching and for improving learning – working with individual students and groups to move them on from where they were at on a daily basis.

Well, the expectations are legal requirements, through ERO. They stated quite clearly that there’s a great lack [of records] in the school, so we’ve got to meet those. (Katy, Waimaka)

There are policies for each curriculum area. There are certain things we have to achieve. For science we have to do a fair test.... We have to give the information to [the senior teacher] or whoever, [to show] whether our children can do a fair test. (Lisa, Arawhata)

[Good assessment] is something that’s practical to me, that I can use, not just doing it for the sake of someone else, to put on file and put it in a box, which happens quite a lot. (Sam, Arawhata)

Arawhata School's report also recommended that the board should consider "aggregating and analysing student achievement information to establish a baseline to measure progress". Although it was not stated why this was important, it seemed that the ERO believed that school-wide assessment data, aggregated and analysed, would inform teaching and learning. This was not addressed in either of the ERO reports, nor was it described by any of the teachers as a useful approach.

However, documentary analysis of two ERO texts, *Assessing Student Achievement* (Education Review Office, 1995a) and *"Barriers to Learning"* (Education Review Office, 1995c), shed some light on these aggregation recommendations. These two reports were produced in 1995 and were available at the time the results reported here were gathered in 1996. In *"Barriers to Learning"* ERO stated that there was no clear definition of such barriers. However, ERO appears to work on the premise, based on the various government policy documents, that there are groups of students, such as male and female students, students of all ethnic groups, and students with different abilities and disabilities, whose learning is impeded by barriers to learning. Therefore, the recommendation in this report is that schools need to aggregate and analyse assessment information in order to identify barriers to learning and achievement for particular students or groups of students. The report stated that instead of using these sorts of evaluative techniques, most schools have interpreted barriers to learning as characteristics of the home and family, poor literacy levels, behaviour management problems, or poor attendance, that is, a student deficit model, where the barriers are attributed to the children rather than the teachers. The conclusion to the report, despite stating that there had been an absence of any definition of the term in the legislation or the policy documents, found that

few boards and their professional staff have fully understood the connections between the concept of barriers to learning and notions of incentives to learning and educational achievement. It is clear that the concept and the responsibilities it implies are difficult for schools. (Education Review Office, 1995c, p. 26)

A lack of any explanation about what actually leads to improved learning – that is, formative assessment – is manifest in this document. After an introduction and comment on the policy framework, the report sets out its findings about school responses to barriers to learning requirements. It notes that most boards are interpreting barriers to learning as exogenous – that is, largely external to and beyond the control of the school. In other words, the report found that schools often located the barriers within the students, their families and their homes, rather than in classrooms. The advice that follows is that schools and teachers should identify barriers that are within the power of the school to do something about. One paragraph notes that

Most good teaching programmes begin with a diagnostic phase during which the teacher assesses the success of prior teaching and establishes how best to proceed in order to encourage the next learning steps. It is the teacher's job to find a way to engage and sustain the interest of all students in the planned learning programme. Teachers need to acknowledge students' backgrounds in planning these programmes but they cannot use them as an excuse for lack of achievement. (Education Review Office, 1995c, p. 8)

The sections that follow on analysing and addressing barriers to learning focus upon collecting comprehensive information about student learning, aggregation of that data and the analysis of it to identify factors that may impede learning and changes that may improve outcomes. While these are important, and the report suggested that schools should examine their “own practices critically in order to ensure they meet the needs of students”, there is an emphasis on school-wide analysis and managerial approaches to improving learning rather than the formative micro-practices occurring during learning and teaching that research has demonstrated make the most impact upon learning (for example, Black & Wiliam, 1998b).

The fact that the ERO had had to interpret the term “barriers to learning” from the implications of government documents could have meant that their interpretation was only as valid as that of the 270 schools they had drawn their study from. My point here is that the meanings of such terms as “barriers to



learning”, “monitor student progress” and “assess student achievement” are open to a range of interpretations. Language is not transparent and meaning is plural and contested (Weedon, 1987). In constructing what they mean by barriers to learning the ERO has included discourses of accountability and equality of educational opportunities and outcomes for groups of students in its definition. Consequently, this meant promoting systems of aggregation and analysis by group as the means of school effectiveness. In contrast, the boards and staff at the schools they reported on interpreted the meaning somewhat differently and acted to address barriers to learning at a more individual or student-centred level.

*Assessing Student Achievement* (Education Review Office, 1995a) focused almost entirely on formal aspects of school-based assessment and records of student achievement when referring to primary schools. Informal modes, such as self-assessment, anecdotal assessment and peer assessment are noted as being commonly used in classrooms but are not addressed in this document, and formative assessment is not mentioned at all. Under the heading “Essential Characteristics of Assessment Information” the report set out what it called general “good practice” statements for assessment (Education Review Office, 1995a, p. 10). After six relatively general principles of assessment, the seventh and eighth stated:

The school should have systems that ensure achievement information can be brought together so that judgements may be made about the achievement standards and progress of groups of students as well as individual students, and achievement information should be gathered and presented in ways which do not give a distorted picture of student achievement. (p. 10)

These give emphasis to summative assessment used for making judgements about groups after teaching. It is up to the schools to decide how the “groups” are defined. It is also assumed that achievement should be measured in the same way, irrespective of the culture of the groups. No mention is made in these good practice statements of using assessment information to inform future teaching or even to address barriers to learning, and no attempt is made to identify barriers to learning first or address these during teaching. Although the purposes

of assessment are stated as providing information to students and parents, informing teaching and learning and informing policy decision makers such as government ministers or boards of trustees, the picture gained from reading both of these ERO National Evaluation Reports is that schools are expected to gather assessment information mainly for the purpose of reporting it. The teachers' interpretations of NAG 1 (ii) and (v) were quite consistent with the accountability messages available to them. Compliance, therefore, appeared to result in parallel systems of assessment carried out to meet the NAGs on one hand and to inform learning on the other.

Both Arawhata and Waimaka Schools in remodelling their school-wide systems, were finding new ways to keep track of both individual and cohort achievement. By "cohort" I mean the achievement and progress of groups of students with something in common. Cohorts at Arawhata School were the year groups whose progress (change over time) was sampled periodically in particular essential learning areas.

We use... sampling, which we do in written language. We gather up the reading levels of the children and report them to the board of trustees. We're gathering up all that data for oral language, written language and mathematics as well. It's very much an accountability thing for the school. We've got sampling in art as you've probably seen. (Julie, Arawhata)

Both schools looked carefully at the results of their six year old cohort each year in terms of literacy as judged using the Six Year Net (Clay, 1979) and Arawhata used a school entry test with every five-year-old new entrant.

Our entry test is planned and prepared to the needs of our particular school. I don't believe the same entry test would be applicable to every school. You have to do it to know your clientele area. It tells us a lot. We do that within the first two weeks at school. It tells us a lot. It's very basic, on language acquisition, early maths knowledge, and their oral language. (Anne, Arawhata)

They used these results to report to the boards of trustees and as evidence of school-wide review, identifying, analysing and addressing the barriers to learning for their ERO audits and reviews. Discussions with Julie, the acting

deputy principal at Arawhata School revealed that her school also used this information to prepare long term school-wide strategic plans for the following year. Julie also described how she kept records to show a picture of the school intake over time, but it was not clear how this picture informed the teachers' classroom practices.

These school-wide efforts to gather achievement information were clearly linked to the accountability discourse of assessment identified through the observations and interviews with teachers. Influenced by the national requirements (the NEGs and the NAGs), the ERO reports that they had received and the current focus on teacher accountability to government and the public at large, these schools were refining what Foucault describes as surveillance mechanisms required as part of panoptism.

As outlined in Chapter Two (pp. 52-3), using the metaphor of panoptism in education helps to explain how surveillance can operate by looking and recording. The administrative and curricular restructuring in the early 1990s required each school to monitor and improve the achievement of each of its students, to record his or her progress and to report on progress to the board of trustees, the community and the ERO regularly so that the school itself, and the teachers within, might be monitored. Through the technologies of writing records, compiling student files of achievement and aggregating the achievement of groups of students, the teachers were both monitoring the performance of the students and, through doing this, disciplining themselves. In attempting to comply with the requirement to monitor, record and report on how students were achieving, and finding this time consuming and stressful, the teachers at these two schools, by undertaking a review of their practices and procedures, were developing new, more efficient technologies to keep their students' learning under surveillance. The clarification and use of very precise and specific systems to monitor student achievement, such as these two schools were in the process of inventing and refining, exemplifies what Foucault (1977) describes as the "examination". Rather than the use of examination as a test, Foucault saw the examination as a process

that places individuals in a field of surveillance [and] situates them in a network of writing; it engages them in a whole mass of documents that capture and fix them. (Foucault, 1977, p. 189)

Foucault thus saw the technology of writing, reading and keeping records as critical to the functioning of power. He saw the functions of this “disciplinary writing” as the formation of a whole series of codes of disciplinary individuality that made it possible to transcribe, by means of homogenisation of the individual, features established by the examination through an educational code of conduct and performance (Foucault, 1977). In addition, disciplinary writing

concerned the correlation of these elements, the accumulation of documents, their seriation, the organisation of comparative fields making it possible to classify, to form categories, to determine averages, to fix norms. (Foucault, 1977, p. 190)

The teachers at the two schools in this study certainly did feel as though they were engaged in a whole mass of documents that captured and fixed them. Apart from the ERO recommendations, a major reason for reviewing the accountability procedures to make them more efficient appeared to be the teachers’ desire to make more of their time and energy available for teaching.

I like to teach but I don’t like to write things down. In my ideal school there’d be less writing for teachers, I think we spend a lot of time planning a unit, planning how you’re going to assess it all. I realise we have to collect data, especially for writing reports at the end of the year, but in the ideal school there wouldn’t be the pressure to collect all the data. (Rose, Arawhata)

It’s not my job to wander around all day with a clipboard in my hand writing notes. I’m there to talk to the children, to develop their academic skills and their social skills as well and I can’t do that if I’ve got a clipboard in my hand and saying “oops, sorry, just a moment everybody I’ve just got to tick my sheet here and here”. (Chris, Arawhata)

So the teachers were taking part in the respecification of their school-wide recording and reporting systems in order to make them more efficient and effective and, to some extent, to find ways to use the same procedures to inform their

teaching and improve learning. The teachers also wanted to clarify the school-wide procedures to alleviate the guilt they felt about how much (or how little) they were recording. Sam, at Arawhata School, described her concerns about the profiles thus:

The reason I don't like the profiles...is that there's more and more assessment [recording] being placed on us and there's still only x amount of time in the day. And I sort of have this guilt thing that if I don't do it [all] then I'm not doing my job. You're being asked to do more and more and more and more and more assessment, and [although] a lot of it is...less practical, it's what we need to assure other people that we are doing our jobs. (Sam, Arawhata)

Sam felt the assessment requirements as surveillance mechanisms by other people to such an extent that they interfered with how she was able to perform as a teacher. She experienced this situation as a dilemma (Woods et al., 1997) in her present job but explained how she had left her previous teaching position due to the overload of assessment demands.

It's getting that balance and still remembering that we're real people with real lives and that we can't do this 24 hours a day. I actually quit my last job because of it. I actually quit. I quit because I had enough. It was the paper pushing that just got to me. I'd absolutely had it. I was sick to death of thinking, oh, I have to go home and put all these little bits of paper in the profiles. I don't see the maths [tests] and the running records so much as assessment because that's what I use to do my planning. That is part of it [teaching] and it's practical. I need to do that. I can't do my next day without having done that. But it's all the profiles and extra things that are there to please other people and not for me.

[Mary] Who are the other people?

Oh, the big ERO thing over your head really. You've got someone, like your team leader, that tells you that you have to do this because someone higher up's told them that ERO are peering over us to check that we're doing our job. And you can never do enough to please ERO. That's the sort of feeling that's out there. One teacher in another school told me that ERO are going to find something wrong, so there might as well be something practical [so not to bother writing it all up]. But [you have to because] you get [pressure put] on you because ERO are

going to come in, they are going to look in your school, they are going to assess you and it's going to be in the paper. If it's bad you don't get kids coming to your school and you're left without a job. And you're letting everyone else in the team down if you haven't done it. You're letting the whole school down. So it's a guilt thing.

[Mary] It's a brave move to come here and take it on again?

It's crazy. I thought I'm never going near a school again. I was sick of the system and the politics, all that sort of thing. But this school is really, really organised and when you're new they tell you heaps. Even so, there's still so much you find out by accident. (Sam, Arawhata)

Sam's obvious distress appeared to be triggered by emotions of guilt and uncertainty due to the changes she felt that she was expected to make to what she recorded about student's learning. Hargreaves (1994) argued that similar reforms in education internationally set guilt traps socially located at the intersection of four specific paths of determination and motivation in teachers' work: "the commitment to goals of care and nurturance, the open-ended nature of the job; the pressures of accountability and intensification; and the persona of perfectionism" (p. 145). Sam said that while she felt comfortable with assessment that was necessary for her planning and teaching she was uncertain about how much she should be recording, and this led to increased record keeping and to feelings of guilt in case she wasn't doing her part for her school in the event of an ERO review. Woods et al. (1997) also found this to be the case in the United Kingdom following restructuring. They note that following the re-formation of the curriculum and assessment demands there was ample testimony about overload on teachers due to increased bureaucracy and to the heightened pressures of competition and accountability. They argued that these developments were changing the nature of teachers' work from one characterised by the problematic but relatively free resolution of dilemmas to one in which teachers began to feel more oppressed by tensions and constraints.

Within formal expectations, teachers have to be both teachers and managers, schools have to be autonomous yet controlled, teachers have to collaborate yet compete, work individually and

collegially, are excluded from policy implementation yet vital for putting reforms in place. (Woods et al., 1997, pp. 12-13)

Sam's description of assessment and recording also raised issues about time and place. She acknowledged that assessment was necessary and that she couldn't do her job without it, but she felt that much of her life was being consumed by the need to keep profile records as exemplified in her comment that she couldn't keep this up 24 hours a day. Hargreaves (1994) argued that in many ways teachers' work outside the classroom was becoming a highly contested region between teachers and administrators in terms of what he calls its front-stage and back-stage properties. His investigation revealed that time was of prime importance to teachers and that their time out of the classroom was subject to colonisation by school administrators for their purposes. Sam described the same problem. Time she had previously devoted to planning for teaching, often using assessment information she identifies as "practical", was now demanded for filling in profiles that she did not need and were primarily to assure the ERO that she was doing her job.

Sam, although most profoundly affected, was not alone in her feelings that precious time was spent recording information for other people. At Arawhata School other teachers questioned whether the detailed profiles would be of real use to anyone.

I question whose going to pick them up and go through them and see what a particular child has achieved in science – whether they know a certain objective. I mean, who is going to look at it and what use is it going to be. And if it's not going to be of any use then why are we wasting our time doing them? I get the profiles at the beginning of the year and... I look at their reading ages to find out where I can start them, and their maths level to find out where to start them, and I look at their social comment. If I've got a concern, I'll just compare it with the teacher last year. But apart from that, I don't look particularly at them. (Chris, Arawhata)

The teaching principal at Waimaka also alluded to a need for balance.

I don't think everything should be assessed. But as a school I suppose we have to decide what it is each year, or each term, that

we are going to assess, and whether in fact we assess each child. What information we send on to other schools and what our requirements are as far as reporting to parents. (Katy, Waimaka)

As these examples demonstrated, teachers seemed very aware of the need to be accountable to meet governmental, Ministry, ERO and school requirements. These were obvious effects of the managerial discourses introduced into self-managing schools in the early 1990s. However, there were also other, less obvious but, I would argue, quite powerful discourses of accountability at work in both of these schools. I turn to these next.

### Accountability to Colleagues

Although the teachers did not state directly that they felt accountable to their colleagues, many of their practices indicated that this was the case. It was assumed in the nature of the job. For example, Rosemary said:

It's my job for a start, isn't it? I am responsible to see where kids are at and what teaching needs to be done. So really I think the whole thing falls on my shoulders. I don't see it as anyone else's responsibility apart from any information that the principal or the DP may want. As far as school-wide reading information may go or trends in reading at our school, like the whole thing falls on me. (Rosemary, Arawhata)

Her statement that she was responsible to see where kids are at and what teaching needs to be done related to the agreement between the teachers that they would monitor student achievement both for their own needs and for school-wide purposes. Doing her "job" implied that other staff members were relying on her to provide certain information and fill in the student profiles they had decided upon as policy. The commitment to "doing her job" may have been related to the collaborative way in which the assessment policies were developed at Arawhata School.

We all sort of shared how we were going to do the profile, the big one that we're collecting. We had meetings on that and discussed how we wanted it set out, we had input into that. (Rose, Arawhata)



Rose was also keen to fit in with what was going on and do her part, although she was a little unsure just what this might entail.

Well, I might even put it [pressure] on myself, I don't know. Like I'm new back here, lots of things have changed. You want to make a good impression, you want to be seen to be doing the job well. I know that in this school there are quite high expectations. It's worth it, you know, my children go here and I'm really pleased.... But we're on the other side of the coin and we're doing the work as well, so I may put pressure on myself, I don't know, that's what I think. We don't know how much other people are collecting. I could find out that I'm doing a fine job and that I'm doing plenty but you're too scared to ask anyone.  
(Rose, Arawhata)

But Rose also saw it as her duty to support her colleagues.

I've got the impression that everyone's far more accountable right through, all the way down the line now. Because [the DP] feels that and [the principal] feels that too. For them to be seen to be doing their job, we have to be seen doing our job and it's just a real filter-along thing now. There were always the inspectors, weren't there? I don't know why I feel that it's intensified but I feel that it has. I think that it's not just in education, it's in all areas now everyone's more accountable.  
(Rose, Arawhata)

Timperley and Robinson (1997) argue the prevalence of a weak version of what they coin "professional accountability" in New Zealand schools. They present survey and case study data to support their belief that the nature of the micro-politics supports this form of accountability. Kogan is quoted as defining accountability as "a condition in which individual role holders are liable to review" (1986, cited in Timperley & Robinson, 1997, p. 25), and their view is that when this definition is applied to the school situation, it is inclusive of accountability to peers, school leaders and managers, as well as to non professional interests such as the local community and governmental agencies. Timperley and Robinson link Strike's (1990, cited in Timperley & Robinson, 1997) two types of democratic control – bureaucratic and communitarian – with accountability to government and the community respectively, but argue that in

tension with both of these democratic sources of control is professionalism. Within professionalism teachers are regarded as a self-regulating and self-disciplined body of professionals who are collectively responsible for assessing the quality of their own practice (Timperley & Robinson, 1997, pp. 158-9). They explain that within this discourse, teachers are accountable to their peers, and the role of school management is to support teachers and provide resources to meet their developmental needs.

Although there was some evidence of a form of professionalism contained in the teachers' stories of assessment and that it was in tension with other discourses of accountability, the examples quoted above indicate that these teachers responded to the hierarchical control of school managers in providing information about learning. However, there was also a sense in which these teachers put pressure on themselves through what Foucault described as the pervasiveness of power in modern society. As introduced earlier, Foucault's concept of disciplinary power explicitly shifts analyses of power from the macro-realm of structures and ideology to the micro-level of bodies (Gore, 1998).

Foucault elaborates on the invisibility and pervasiveness of this type of power in modern society through the use of various techniques including surveillance, normalisation, individualisation, exclusion, and (self-) regulation (Gore, 1998). While the teachers did not report that they were accountable to their colleagues, and may not necessarily have been aware that they were, the mechanism of self-management appears to have increased peer and self-surveillance in order to ensure the implementation of school assessment policies.

### Accountability to Students and Parents

Teachers also indicated that they were accountable to the students and their parents in terms of assessment. Strike's (1990, in Timperley & Robinson, 1997) second type of democracy, communitarian democracy, includes teachers, parents and students, and has been given greater emphasis as a source of legitimate authority in New Zealand schools through boards of trustees. Many responses indicated a duty to be accountable to the board of trustees.

We gather up the reading levels of the children which are reported to the board of trustees. (Julie, Arawhata)

That's because their [the board of trustees] job is to put all those things [policies] in place really. And they're more answerable to the powers that be. I'm answerable to them but they are answerable to ERO. (Sam, Arawhata)

But there was no indication that the teachers saw this as being accountable to students and their parents. Accountability to parents was referred to in many statements the teachers made about reporting on progress of individual students, such as

good assessment can be understood by other people, something that can be translated for parents because parents like to know what's going on as well. (Rosie, Waimaka)

At parent interviews I'm more than happy to have parents view exactly what I've put down about their child. To me it's more their business than mine. It's their child, so anything I do with reporting to parents is there for parents to view. And having it in this format [in a roll-type book] I find it quite easy to be able to show them their child without anybody else's. It's easy to block off their names, but they can see exactly how I think their child is doing. (Anne, Arawhata)

Some parents are very achievement oriented and they want their children to succeed. So I sort of think well, I'll give them an exercise book for homework and it's got basic facts for maths, writing stories, and spelling activities in it. Some parents have [been pleased] and it will help me when I see what they've done. (Luke, Waimaka)

From these examples, and the other responses I received, it was clear that accountability to parents for these teachers meant keeping them informed about their own children's learning. This was consistent with findings published by the ERO in 1999, stating that

Teachers in many schools have records of a large number of assessments of learning objectives or statements of outcome for many areas of the curriculum for each child. Often the only use made of those records is to provide information that helps teachers report to parents. (Education Review Office, 1999, p. 34)

This is unsurprising as traditionally this has been the major use made of summative information collected about individual students. Keeping School Records, the system of record keeping scheduled for complete implementation in 1989 before it was overtaken by the administrative and curriculum restructuring of that time, held this to be a major purpose of a school system of assessment and evaluation (Department of Education, 1989b). The booklet for teachers stated that the new (Keeping School Records) system should “provide useful, ongoing information, which will help teachers to review students’ progress, plan for future work, and make useful and objective comments to parents about their children’s learning and development” (Department of Education, 1989b, p. 4). It also provided a section on how to make reports to parents objective which included clear guidelines about how to describe what a student had covered, what he/she could do, and the criteria by which he or she was assessed with specific comments (pp. 21-3). As I have argued in Chapter Two, all through the 1980s, teachers throughout New Zealand were extensively involved in the development of the Keeping School Records system and had taken on board and, at times, led the call for new criterion-referenced discourses of assessment. What had changed in the post-1989 environment was that now schools were self-managing, the prime use of this summative information was seen by agencies such as the ERO and the Ministry of Education as a source of data with which to judge school’s performance. The information from this case study and the ERO education evaluation report (Education Review Office, 1999) suggests that there were tensions between the different accountability discourses in the mid to late 1990s.

### Formative Discourses of Assessment

Accountability discourses and the tensions between them were not the only forms of surveillance that were evident from my early contacts with the schools. Many of the teachers were wrestling with other assessment dilemmas, related more to classroom practice than to accountability demands. From my observation and the interview data, I was able to identify a further set of assessment discourses

related to teaching and learning as opposed to accountability. Because these discourses related to how teachers recognised and responded to student learning in order to enhance learning during teaching (Bell & Cowie, 1997), I labelled them “formative discourses of assessment”.

The Ministry of Education defined formative assessment as

A range of formal and informal assessment procedures (for example, the monitoring of children’s writing development, anecdotal records, and observation) undertaken by teachers in the classroom as an integral part of the normal teaching and learning process in order to modify and enhance learning and understanding. (Ministry of Education, 1994, p. 48)

Similar definitions are contained in the assessment literature (Black, 1998; Gipps, 1994a), but, as Harlen (1998) points out, formative assessment does not, unfortunately, enjoy a widely recognised and agreed meaning. It is often confused with such terms as “teacher assessment” (of students), “ongoing assessment”, “continuous assessment” and “diagnostic assessment”. In reality, Harlen suggested, these terms often have more to do with assessment in the form of a series of short summative tests than with a form of assessment that is an integral part of the teaching and learning process. As explained in Chapter Four, her remedy was to use “the more clumsy phrase [of] assessment for formative purposes when referring to assessment that is really used with the intention of supporting learning and, indeed, actually helps learning” (Harlen, 1998, pp. 2-3). In my interactions at Arawhata and Waimaka Schools the teachers never used the term “formative”. Several referred to “child-centred” assessment as the most valuable assessment they used. They described child-centred strategies as those that were aimed at finding out where children were “at” (not where they had moved from and to) so they could assist them to make progress, in contrast to those that were undertaken and recorded for reporting. I have interpreted these “finding out where children are at in order to move them along” strategies as formative ones.

Within the set of formative assessment discourses there were variations, however. Some teachers relied quite significantly on a range of informal tests to

make decisions about learning and teaching and move it along. Others were more focused on conversational and observational techniques. A few of the teachers talked about introducing student self-assessment within their programmes. Each of these discourses of formative assessment is addressed in more detail below.

### Informal Testing

Teachers at both schools (but mainly from Year 3 classes up) used a range of informal tests to assist with the learning process. Maths pre and post-tests were the most prevalent. Many felt that the only way to find out about what children knew and could do in mathematics was to test.

Other than that [self-evaluation] just the usual testing, for example in maths, is sometimes the only way you're going to get knowledge of what the children know – it's a straight out test, before and after, which our children don't seem to mind doing which is really neat. I use assessment primarily for planning more than anything...like a pre-test for maths – what ideas do they have, what ideas don't they have at all, and then, OK where do we need to focus? Then you have a similar test at the end to see what's happened. So planning's a big one, but we group according to ability a lot of the time. (Lisa, Arawhata)

I'm a bit old fashioned and I like an oral or a written test, especially in mathematics. I mean you can't really beat them. I use the running record [for reading], we have spelling tests every week, vocab tests really. We do maths pre- and post-tests and for science and social studies we do an oral discussion to find out what they know before we carry on with [the unit]. Afterwards would be activity based, like worms, they had to draw them and write down what the parts are. I'm not into a closed book test...I mean, they've had...work-sheets and stuff. I also see it [their learning] because we chat...like when a worm was found when we were out doing PE, they were able to say what part was which. (Luke, Waimaka)

[In maths] I do pre-tests to find out on a topic what they already know. One or two children don't have to do lots of things about that [because they do well on the pre-tests]. Then I do exactly the same test after my unit so see how far along the track they are, whether they've picked up anything from my teaching. And I do basic facts tests at the beginning of the year and a few times

during the year to see whether [they are making progress]. (Viv, Waimaka)

While several of the teachers were happy to use the sorts of non-standardised tests referred to in the quotations above, there seemed to be less value placed on standardised norm-referenced tests for formative purposes. For example, Chris said that it was school policy to use Progressive Achievement Tests (PATs) and that the results were aggregated and graphed by the deputy principal. But, he stated,

it comes back to [the fact that] some children score poorly on PATs when they are actually the ones that when you talk to them they are able to do it. I actually read some of the children the maths PATs because it's about maths not reading. We do use the PATs, they're all recorded but not [used for teaching]. (Chris, Arawhata)

I don't feel they're [PATs] accurate. It's very hard to trust them. Maybe that will come with experience but I'm still not quite sure what they're on about. It also puts the child in a box. I think teachers like boxes but maybe they are not always true. So the child goes through the year being seen as good at maths [and treated that way] but in effect they may not be as good as the PAT has reflected. (Rosemary, Arawhata)

From this type of information it seemed that many of the teachers were not clear about the use of the PATs and were suspicious of norm-referenced tests that had the potential to label children as performing poorly when the teachers believed they could perform better than the test showed. Although in the PAT manual it states that the primary functions of the tests is "to assist classroom teachers in determining the levels of development attained by their students in the basic skills...to enable the teachers to make decisions about the kinds of teaching materials, methods and programmes most suitable for their students" (Reid & Elley, 1991, p. 3), where they were used, the information often appeared to be recorded for reporting purposes and regarded suspiciously by the teachers. In fact there was some antagonism towards the fact that the teachers were asked to administer these tests. It appeared that, in these two self-managing schools at

least, these tests, while not used by the teachers formatively, were meeting the managerial demands for data about whole groups of students.

The information from the teachers indicated that although they were suspicious about norm-referenced test information because they knew more about their students than the tests seemed to reveal, they used several teacher-made and less formal tests (like maths pre-tests) because they were manageable. For example, when commenting on a questionnaire that he had designed to assess a science unit, Chris said that he

went through it and decided whether the children had actually grasped the concepts that we had tried to teach them or not. I like that because you can do it as a whole class. I mean, it would be ideal to sit down and have a few questions you could just ask each child but you haven't got the time. In that manner you'd get to know exactly what the children have learnt but there's no possible way that that can be done. (Chris, Arawhata)

In this sense manageability appeared to be a way of balancing the best possible information against the means for getting it. It would seem from this explanation that as there is no time to do it all individually, this teacher used an informal test with the whole class (or a group) to get the best information in the circumstances. In this way the written response provided by each child could be used as the material for each case without having to study each child's knowledge in a one-on-one manner.

The only situations where teachers appeared to use one-on-one tests were (reading) running records and diagnostic tests in the junior school, such as the Record of Oral Language (Clay, Gill, Glynn, & Salmon, 1983), and the Six Year Net (Clay, 1993). However, I did not get enough information about these at this stage to know whether the running records were analysed for strengths and needs, or just used to calculate a reading age with which to describe progress.

### Continuous Assessment

While some teachers relied to an extent on tests with groups of students, others did find manageable ways to talk with and observe individuals as their main



form of useful assessment. This was particularly (though not exclusively) the case in the New Entrant/Year 1/2 classes. These teachers explained assessment as an integral part of teaching, including diagnostic testing, informal running records, observing children as they worked and talking with them about what they were doing.

You're doing it all the time. You're observing children and you're acting on what you see immediately. I've found that really easy to do with having a small class up until now. Easier than when you've got 37 and struggling away to get around everybody. I also try not to do so much at once. I came from a school where we were encouraged to analyse everybody and everything all the time and I don't think it's necessary to do that. (Marvin, Waimaka)

When I do my small group work I'm observing and watching what they're doing. Whether they're using their dictionary confidently or whether they're looking over at their neighbour. So those little, small observations I find quite helpful. (Rosie, Waimaka)

Observation. That's my main thing, observation on the go the whole time. While they're all participating. I can see whether they're confident, for example, conversing with each other and me, that their speech is correct or volume adequate, or if they can organise their thoughts. I know that they're within the skills that those children need to acquire and so I can tell [if they can do it]. And, if it's not happening, then I can work on it to get them more confident that way. (Anne, Arawhata)

Asked if they remembered the information so they could act on it later, these teachers described various strategies for recording information for formative purposes. For example, Marvin said she records what is useful in her roll book and she records what children say and how their ideas have changed. Anne also had a roll book and explained:

It's with me, nice and close to me whenever I want it. Everything [is in it], and notes, and I find now that I know the document so well and the things that I am doing, that as the children are doing their work I can quickly just note down what's happening or what they need. Some have not participated much in this unit so I know for next time to encourage them. I have it

on hand. I couldn't work without it. Just like with my weekly workplan. I need to refer back every so often. (Anne, Arawhata)

These teachers did not necessarily have small classes (one had over 30 children) but seemed to have a range of organisational strategies that allowed them to individualise their assessment. What I mean by this is that they were very clear about what they were looking for as signs of learning and organised their teaching to facilitate the recognition of these indicators. For example, Anne told me

A lot of the assessment is day-to-day and week-to-week. It's not once a term, it's ongoing all the time. I've got sheets for the alphabet and basic sight words. I teach in context but I do test for alphabet and basic sight words. I'm doing my running records all the time. We compare handwriting and written language once a term...to monitor progress from term to term but mostly there's no surprises in what comes out as you've been working with it all term anyhow. Maths is day-to-day which I have built into the same sort of system [focusing on particular objectives]. (Anne, Arawhata)

The assessments that these teachers described were examples of how surveillance and intervention effects learning; that is, Anne is looking for change over time. Through breaking down multiplicities into simple units (Foucault, 1977) objectives, and watching the development of each for each child, these teachers were able to intervene and control the learning as it occurred. They did this through being well organised in terms of their routines and timetables (Foucault, 1977), and by having developed orderly ways of identifying, classifying and recording learning behaviour. These techniques are clear examples of how Foucault's concept of disciplinary power works in practice (Foucault, 1980; Gore, 1998). When teachers closely monitor and observe students in this way, surveillance singles out individuals, regulates behaviour (of both students and teachers) and enables comparisons to be made. As Bell and Cowie (1997), among others, have explained, interactive formative assessment is most powerful in generating learning precisely because the teacher can, through close surveillance, notice what the student does or does not seem to know, recognise its significance for the student's personal, social or academic development, and respond

appropriately to change the behaviour in some way (Bell & Cowie, 1997; Pryor & Torrance, 1996). In this sense, rather than as some grim conspiracy theory as some seem to see it, Foucault's notion of disciplinary power can be seen as productive. As Foucault stated

A relation of surveillance, defined and regulated, is inscribed at the heart of teaching, not as an additional or adjacent part, but as a mechanism that is inherent to it and which increases its efficiency. (Foucault, 1977, p. 176)

When asked where these techniques of individual observation, interaction and intervention came from, and how they were developed, one teacher described a personal journey of experimentation.

I've formed this whole booklet [learning outcomes she is looking for in each curriculum area] after all my years of teaching. At the moment [I feel] I've got the recipe right because it is so manageable. It's completely child-centred as far as the stages the children work through. I can pick it up at parent interviews and feel that it's open to them to see their child as well. It's not too high-powered theory or anything, very practical and very useful. But it's taken several years to get it to this stage. At the moment I haven't got any improvements. Most years I've made improvements, improvements, improvements. So far this year I haven't come up with any "oh gee, perhaps I could try it a different way". It has got to this stage after about five years of being. (Anne, Arawhata)

Anne uses terms such as "child-centred" and "stages" in her explanation that indicate she holds progressive ideas about the developmental nature of education. Anne trained to be a teacher and entered teaching at a time when stage theory and progressive discourses of education were predominant. As explained in Chapter Two, from the 1960s Piaget's stage theory of development heavily influenced teacher education programmes, and the textbooks used at that time were underpinned by strong developmental discourses. Walkerdine (1984) used Foucauldian notions of power to explain how, from the 1960s, developmental psychology and child-centred pedagogy together produced the kind of child-centred classrooms Anne was speaking of.

It is possible to disentangle a complex web of related practices and apparatuses which together produce the possibility and effectivity of the child-centred pedagogy. Central to these is a system for the classification, observation monitoring, promotion and facilitation of the development of a variety of aspects of individual psychological capabilities. In apparatuses and practices such as these it is axiomatic that there must exist a set of observational and empirically verifiable facts of child development. Central to the practice therefore is the production of development as pedagogy. By this I mean that development is produced as an object of classification, of schooling, within these practices themselves, made possible by the apparatuses which (among many others) I have singled out: record cards, teacher training, classroom layout and so forth.... It is in this sense that developmental psychology and the child-centred pedagogy form a couple: the apparatuses of the pedagogy are no mere application but a site of production in their own right. (Walkerdine, 1984, p. 162)

Traces of these discourses suggest that Anne has built her practices with developmental theory as a predominant factor. But she also emphasised the practical, manageable features of her approach.

Anne's journey, however, was not limited to trying things out within the four walls of her classroom. She had shared her ideas with teachers in her school and internationally, refining her own techniques as she went.

They are all things that I need to know before I can encourage them on to the next step. Give me another 12 months and I might have a different answer but at this stage in seeing what I did over there in Hawaii.... They're just so far behind. I came back here and thought "gee, these kids here are brilliant", children after a year at school (in the USA) are barely reaching the standards that we get after three months, and their children at standard four level are probably at the end of our J3 level. They're really enthused about the changes they can make. It's exciting. It made me look and see what we've got around here, we're privileged with the resources, the equipment. (Anne, Arawhata)

In this explanation Anne demonstrated how important having defined levels has become within a developmental model. In a linear/hierarchical model, each child's performances can be individuated and classified against a

predetermined progression, for example, Anne's use of Standard Four (10/11-year-olds) and J3 (7-year-olds) as levels in her comparisons between the children she had worked with in New Zealand and Hawaii. As a teacher of many years experience, she did not require the outcomes delineated in a curriculum document but had constructed for herself a set of "steps" that she saw the children as moving through as they made progress. These were the headings she used in her roll book and shared with the teachers she worked with.

What I've got here I have thoroughly shared with my team of three teachers. They are all doing the same as I am. Simply because they've looked at it and they've said, can we copy it? That's fine if it's going to work for them like it works for me. And several other teachers in J2 and J3 have also copied it, so if it's a help to anybody, they are welcome. (Anne, Arawhata)

This sharing within schools and between schools, and even between countries, influenced the way the teachers carried out assessment in their classrooms and confirmed for Anne that her approach worked.

Marvin, at Waimaka, attributed one source of her assessment practice to professional development courses she had been involved in but added that she was involved in a struggle to use assessment within units rather than checklisting at the end.

After doing last year's course I now build assessment tasks into my units, if you like, instead of putting on an add-on at the end. That's something I feel is very useful, but I still can't get away from checklisting. When I think about it afterwards I wonder what is the purpose of doing it. So I'm trying to ween myself off doing checklists for everything and really just setting myself priorities on whatever things I want to assess. So I think teachers are their own worst enemies in some ways. By nature they tend to grab every idea that they can get their hands on. I think we've actually made a rod for our own backs by trying to achieve all these things and probably only a few of them are really necessary. (Marvin, Waimaka)

Checklists and clipboards appeared regularly in the responses of teachers in these early interviews. Keeping in mind Walkerdine's argument about the central nature of practices, such as "record cards... teaching notes, work-cards, school and

classroom organization [being] no mere application [of pedagogy] but a site of production in their own right" (Walkerdine, 1984, p. 162), there appeared to be value in exploring these in further stages of the research. Marvin also went on to explain how her assessment ideas were influenced by her past experiences.

I came from a school where we were encouraged to analyse everybody and everything all the time and I don't think it's necessary to do that. I think back to what I was doing when I first started teaching. We had a system of evaluating there that I found very useful. It was very simple and it didn't involve checklists and things like that. Every two weeks, fortnightly, we wrote an evaluation, if you like, of the two weeks' programme. And you built into that comments of how the children were doing. But instead of having checklists of all the individual children, you would say, right, there's a group of children who are coping with this and doing this and there is a group of children who have needs in this area. I found it a much more manageable thing to do.... Within it you could actually identify the needs and it was really only once a fortnight that made you really focus, but within that fortnight we knew we were going to be doing this so you'd home in on a few things. But we didn't have roll books and checklists and clipboards all the time. It was really valuable. So I'd like to go back to something like that. (Marvin, Waimaka)

Although all of these teachers preferred to use one-on-one assessment techniques as part of their ongoing programme, there were times when they also used group or class techniques, notably when they were taking term samples of progress in such aspects of the curriculum as handwriting or written language. In all of these cases they indicated that it was school policy to take these periodic samples.

We all know that written language samples are once a term, handwriting once a term...children's profiles filled in twice a year. (Anne, Arawhata)

We do have an existing policy that's quite specific... the personal profiles. I'm using a scrapbook (for samples). (Marvin, Waimaka)

(The portfolio) is supposed to be kept regularly up to date. I take a handwriting sample every term, and a writing sample and it's quite neat to be able to see the progress. (Rosie, Waimaka)

In this way the testing of groups discourse (originating from school policy) existed alongside the continuous assessment in these classrooms without dominating it. Even at Waimaka School, where all the teachers explained that the policies were in flux, there were unwritten expectations about assessment. These were operationalised by the principal when she checked teachers' records to ensure they were assessing the national curriculum. The following statements from Rosie at Waimaka demonstrated these unwritten expectations.

[There's] not a lot of direction at the moment. With the principal changing we've been fairly much left to our own devices. The only direction is to fill out the pink cards.... And to be able to do an oral report to parents.... And Katy [the principal] checks our roll books, or our assessment records.... Later in the year we're going to have a written report.

[Mary] And is the portfolio an expectation?

Oh yes, that has to be kept up to date regularly. (Rosie, Waimaka)

What seems to be invisible in these discourses is how the teachers utilised the information they collected. What did they do with the information they collected in roll books, on checklists and with their clipboards? While assessment seemed obviously linked by these teachers to managerialist exhortations to collect data, their use was not clear at this stage of the investigation.

### Student Self-assessment

A few of the teachers explained how they had involved the children in setting their own goals and assessing their own progress.

I gave the children a piece of paper and said, right, I want to know something that you can do and I would like to know something that you would like to be able to do for yourself over the next term. And they came up with really neat things, like one of my children came up with "I can trust myself" which I thought was brilliant, and then "I want to be able to count in fives to 100" or "I want to be able to make friends in the playground" or "I want to be able to make one more friend this

term". So for me it was really good because the children actually told me what they wanted.... It came straight from the children, it was informal, it was simple. But the parents were really interested to know, too. (Lisa, Arawhata)

I gave them the opportunity...for the children to say how they thought they worked during the unit. So I'm basically putting the onus back on them...to get them thinking about what they're learning. (Chris, Arawhata)

I introduced them (little tests) as a challenge to monitor their own progress. For some children it's just not working. Some think it's pretty neat because, if you like, they're doing tests like the big kids, but most of them, especially the Year 2s, they're not confident and they think "it's too hard" and just don't bother. (Rosie, Waimaka)

While these descriptions demonstrate that they weren't all happy with the results, teachers were experimenting with self-assessment ideas that they had heard or read about.

## **Teachers at the Intersection of Multiple and Competing Discourses**

The findings from this initial stage of the investigation suggested that, rather than the competing discourses of assessment identified as binaries by writers such as Black and Wiliam (1998a), Crooks (1988), Nisbet (1993) and Willis (1994), there were multiple competing discourses impacting on teachers' practices. To explain this a little further, Crooks (1988) drew attention to tensions between the summative and formative functions of assessment, as did Black and Wiliam (1998a). As reported earlier in Chapter Four, Black (1998) explains that when teachers have to be involved with both formative and summative assessment, there is bound to be tension between the two roles. He argues that it is, therefore, the teacher who has to hold the boundary between the different requirements of the formative and summative assessment roles, and indicates that pressures and constraints that originate outside the classroom can drive teachers to move towards



more summative practices at the expense of more private formative ones (Black, 1998).

Nisbet (1993) suggested that the tensions between assessment for accountability and the promotion of learning needed to be reconciled (see Chapter Four), and Willis (1994) identified similar tensions between what she termed technical and educative discourses of assessment. In each case, the tensions are binary: formative/summative; private/public; high-stakes/low-stakes. In these explanations of assessment tensions the teacher has to “hold the boundary” (Black, 1998, p. 120) or find a balance between two roles.

My analysis here, however, identified multiple discourses. While, clearly, these can be grouped somewhat under the two general categories of the summative and formative roles of assessment, I believe that the situation for teachers is more complex than holding a boundary between two roles. In contrast, these teachers appeared to me to lie at intersecting nodes of the varied discourses (Weedon, 1987), weighing up their priorities in each circumstance and acting in light of what appeared to them to be the most pressing or powerful discourse at the time.

The evidence from my observations and interviews was that, in most cases, the teachers responded to the most powerful discourses within their institution. Weedon (1987) asserts that the most powerful discourses in our society have firm institutional bases. In fact she argues that in order to be effective and powerful, a discourse needs a material base in established social institutions and practices.

Yet these institutional locations themselves are sites of contest, and the dominant discourses governing the organisation and practices of social institutions are under constant challenge.  
(Weedon, 1987, p. 109)

The implications of such contest and challenge, even for the school itself as an institution, meant that teachers were wrestling with the impact of these discourses on a daily basis in their work. As I have documented above, the teachers at both Arawhata and Waimaka Schools were attempting to rationalise their school-wide systems of assessment, recording and reporting so that they could focus more time and effort on formative assessment practices. But even

within these formative practices there was struggle involved for teachers in working out which to use. Chris acknowledged this when he stated

I struggle with assessment anyway. Good assessment would be providing you with a means to take the child on to a next stage. Having accurate data on where that child is in relation to your objective set at the beginning of a unit. I guess in evaluating your unit you can decide whether you set a good assessment task or not. But, I also think assessment needs to be varied and different because different children will respond in different ways to different assessment procedures. So it's hard to say, what is good for some children in assessing their work may not work with other children. (Chris, Arawhata)

In this example the teacher can be seen as demonstrating how multiple discourses of formative assessment are competing for his attention. He knows that he wants an assessment device that he can use at the beginning (or end) of his unit of work to tell him how all, and each, of the children can perform in relation to the objective(s) he has set, presumably from the national achievement objectives. But he also knows that the information that one assessment task will give him might be more or less valid or fair for each of his students. In addition to this observation, a little later in the interview, Chris acknowledges that there are other considerations besides measuring performance in relation to the objective he has set and the fairness of using particular assessment tasks. He justifies the use of a questionnaire for assessing a science unit because he can use it with a whole class, saving the time it would take to discuss things with each child individually.

Discourses of assessment, derived from NAG 1 (ii) and (v), impact on the need for Chris to gather information about how each individual performs in relation to the objectives, but the time available does not allow him to do this. So he has to come up with methods that allow him to obtain documented individual case material. Later in the interview Chris states

I spent time on a science initiative last year, run by the ministry. And, in that science initiative, we looked at assessment and just the various ways that you could assess the children. So one of my goals this year is to develop my science units and a variety of teaching methods, and a variety of assessment procedures within that. We've done one science unit...and we used a questionnaire type thing where the children wrote down items. The next one

we might focus on writing a letter home, or doing oral presentations. (Chris, Arawhata)

This statement implies other discourses, perhaps originating from current professional development contracts, have provided Chris with several ways that he can gather these individual documentations. Still other discourses, however, about the purposes of assessment, or perhaps the use to which he will put the information he collects, compete for his attention

But then it comes back to my last question, my last statement of really questioning the need to assess the children in science and social studies and other curriculum areas. I think as long as you're providing the experience, and it's a quality experience, then, anyway,...I'm happy with the way it's going. (Chris, Arawhata)

In the way he leaves his own question unanswered, Chris demonstrated that he was weighing up the possibilities and the reasons for his own teaching and assessment actions. It is interesting that there is very little visibility of group assessments or group performances in these assessment explanations. The prevailing view appears to have a sharpened emphasis upon the individual. Often, culturally appropriate views of assessment such as the assessment of group performances appear to be excluded from these pressures to collect individual data despite the presence of a range of cultures being in evidence within the school's statistical information.

## **Summary and Comment**

Seeing the teachers' situation as the intersection of multiple, competing discourses of assessment facilitated understanding, from the teachers' point of view, why assessment was so problematic. An outside-in view, say from a Ministry, an ERO or an academic perspective, positions the knower (the policy maker, the ERO reviewer or academic) differently at the intersection of assessment discourses. The institutional discourses impacting on these people, while still multiple and competing, may be less shrill due to their priority purposes for

assessment. As I have attempted to show by analysing two of the National Evaluation Reports earlier in this chapter, the priority for the ERO in 1995 appeared to be to increase the effectiveness of schools through the aggregation of summative assessment of individuals. For the Ministry the priority seemed to be for teachers to monitor, assess, record and report achievement against the national achievement objectives. Academics and advisers will have had varied priorities. However, on the inside, at the chalkface so to speak, all of these discourses, as well as the school's own policies, were competing for each teacher's attention.

The evidence from this first round of interviews, together with theoretical notions of power as pervasive and productive, situate assessment at the very heart of teaching through the use of such strategies as the monitoring of individuals through surveillance, individuation, classification and normalisation. In these two schools at least, there appeared to be contestation between the managerial discourses which came "down the conduit" (Clandinin & Connelly, 1996, p. 11) as a prescription of what should happen and the teachers' personal professional knowledge.

While the teachers' and school stories of multiple, competing discourses assisted me to understand to some extent where they were coming from, I was interested to know whether I had heard their voices accurately and interpreted their stories coherently, and I had further questions about the ways in which the teachers believed assessment enhanced learning. Furthermore, I was curious to find out whether teachers at other schools recognised similar discourses of assessment and how they responded to them. To investigate these questions, I continued to observe developments at Arawhata and Waimaka Schools during 1996, I interviewed a national sample of teachers during the reporting and recording action research I carried out for the Ministry of Education in 1997 (M. Hill, 1998a) and I published some papers on my results in order to prompt feedback (M. Hill, 1997a, 1997b). Chapter Seven relates these extended and new stories of classroom assessment.

## **CHAPTER SEVEN**

# **COMPETING DISCOURSES AND TEACHERS' ASSESSMENT PRACTICES**

Discourses make up practical “grids of specification” for diagramming, classifying and categorising the subject in the social. These grids are put to work in institutions in ways which generate self-surveillance, wherein the subject internalises the disciplinary and cultural gaze as his or her own. The effect is one of self-colonisation, where the subject takes on “responsibility” for monitoring her or his morality, discourse, and body. (Luke, 1992, p. 7)

The success of disciplinary power derives no doubt from the use of simple instruments; hierarchical observation, normalising judgement and their combination in a procedure that is specific to it, the examination. (Foucault, 1977, p. 170)

As outlined in Chapter Six, the teachers at Arawhata and Waimaka Schools spent a great deal of time in 1996 reviewing and changing their assessment, recording and reporting systems. By the end of the third term both schools had made decisions about the changes they were going to make to their assessment policies and had plans of action settled for implementation in 1997. They used the last school term to put these decisions into action.

At Arawhata School, the acting principal, Brian, purchased spiral bound exercise books containing five different coloured sections of pages for the teachers to use as data books for their classroom assessment recording in 1997. He also met regularly with the assessment team who had been steering the review and, with their help, drew up a draft assessment handbook outlining the decisions that the staff had made about assessment procedures. During term four 1996, teachers met in their syndicate groups<sup>14</sup> to decide how they would use the data books in

---

<sup>14</sup> Syndicates were groups of teachers teaching at specific year levels. Each syndicate had a senior teacher responsible for the syndicate's functioning. These syndicate leaders performed management functions that linked senior management decisions with classroom practice.

1997, and some sought the assistance of other teachers on the staff to assist them to do this.

At Waimaka School, the teachers came to an agreement about what would be contained in the school record card that they wanted. During term four 1996, Katy, the principal, liaised with a school management adviser to have the card formatted. Over the Christmas break she approached various printing firms to have the card printed, but due to its foldout format found that it was going to be very expensive to produce. In the meantime, the teachers at Waimaka School carried on with their individual assessment and record-keeping systems, and implemented the student portfolio system of work samples described in Chapter Six.

As outlined in Chapter Six, conceptualising the teachers' situations as the intersection of multiple, competing discourses of assessment helped me to explain why assessment was so problematic for the teachers in these two schools. As a result, my interest turned to how the teachers in these schools, and schools nationally, had responded to these multiple competing discourses in the ways they went about their classroom assessment practice. This chapter explores how the teachers at the two Waikato case study schools, as well as 20 teachers from schools throughout New Zealand, responded to the multiple competing discourses of assessment.

## **Teachers' Responses to Multiple Competing Discourses of Assessment**

As well as analysing the data from the first two rounds of interviewing in terms of the emerging assessment discourses, I developed codes and categories to search for possible patterns in teachers' assessment practices (Charmaz, 1995; Gore, 1998) as described in Chapter Five. It seemed important to analyse teachers' practices of assessment in detail in order to understand the effect of the assessment discourses on their practice. As explained in Chapter Four, Black and Wiliam (1998a), in their major literature review about raising standards through

classroom assessment, argued that there is clear evidence that formative assessment raises standards and that the research literature indicates how to improve formative assessment (p. 2). Furthermore, they elaborate on problems and shortcomings of classroom assessment practice, including how teacher testing can neglect quality in relation to learning, how giving marks, grading and encouraging competition is often at the expense of motivation to learn, and how the managerial use of assessment and recording is often at the expense of learning. Therefore, the first two rounds of interviews (already described in Chapter Six) were re-examined with particular attention to the detail of the teachers' daily assessment practices. Using the NUD\*IST software programme, the teachers' responses to each question were examined line by line and coded according to:

- Whether the practices used by each teacher were mainly used for formative or summative purposes;
- Whether assessment criteria were teacher generated or came directly from the national curriculum achievement objectives and/or specific learning outcomes;
- If, and how, the teachers used the assessment information;
- What sort of evidence of learning was collected;
- How, when and for whom the information was recorded;
- If the teachers relied more on memory than recording to store their assessment information.

Such careful scrutiny of the data was also indicated by Foucault's advice that analysing what happens in the exercise of power relations is "flat and empirical" (Foucault, 1982, p. 217). By this Foucault meant that there should be a micro-level focus on everyday practices of surveillance that he saw as being at the heart of the practice of teaching.

The meticulousness of the regulations, the fussiness of the inspections, the supervision of the smallest fragment of life, and of the body will soon provide, in the context of the school...a laicized content, an economic or technical rationality for this mystical calculus of the infinitesimal and the infinite. (Foucault, 1977, p. 140).

In bringing about reform it is necessary to know in detail “what what we do does” (Blacker, 1998) in order to begin to address adequately how things might be done differently (Gore, 1998, p. 249). In contrast to some studies of assessment that have used large-scale questionnaires (for example, Wylie, 1994) and single rounds of interviewing (for example, Dixon, 1999; Thrupp, Harold, Mansell, & Hawksworth, 2000), this micro-level analysis revealed that while some of the teachers reported being very systematic in assessing, recording and reporting assessment information, others appeared to be more intuitive in their approach. Interestingly, in the first round of interviewing, most of the teachers described methods of recording assessment information as their assessment practice. For example, in response to a question about which assessment activities she found most useful, Anne told me that

What I’ve done is set up a book. That’s all my recording and I find that all my objectives...are child-centred.... So I’ve got a focus which becomes part of my weekly planning. (Anne, Arawhata)

And Marvin told me

Well the thing I am really pleased with is recording the children’s comments.... At the beginning of the unit I record a comment about what they know about and then again at the end. I might have some criteria in my head and I’ll say well, look, I can see that this child has actually achieved that through what they said. It often sort of jumps out at you. I write it down in a roll book. (Marvin, Waimaka)

Each of the schools had assessment guidelines that, to some extent, shaped the way in which the teachers recorded the assessment information. It was these guidelines that the teachers in each school had been working on together to improve during the first year of the study. For example, at Waimaka, each teacher had some sort of portfolio (be it a folder or a scrapbook) for each child in which examples of their work were collected and displayed. At Arawhata School information about each child was filed in the individual profiles and, by 1997, each teacher had been issued with a book to keep track of their students’ learning.



But within these school guidelines there was a great deal of scope for individual ways of keeping records. For example, in the roll books, some teachers wrote the names of all their children down the page at the extreme left-hand edge and, across the top, wrote each of the achievement objectives they were including in their teaching units direct from the national curriculum documents. In contrast, while still recording children's names down the side, others wrote specific learning outcomes related to the achievement objectives that they believed they were covering across the top. Another group of teachers simply noted the activity and whether each child had "visited", "met" or "exceeded" their expectations for children at that class level. In many cases the information was recorded against each name as a system of ticks, triangles, letters or colours to indicate coverage and attainment of each outcome, activity or objective.

Although, at first glance, such coding systems appeared similar, there were important differences. In some cases the codes were used mainly to gather evidence of coverage of the teaching objectives and to record when mastery had been achieved. In contrast, other teachers reported using this type of recording as a working shorthand that allowed them to make important decisions about next teaching steps for individual students or groups.

Some of the teachers at the case study schools had very little information recorded at all. These teachers were meeting school requirements to keep profiles and portfolios of evidence about student learning mainly by formally collecting samples for the folder/scrapbook about when it was due or by testing periodically. In the intervening period they tended to keep track of students mentally. For example, Katy told me that she thought

Assessment is by observation, talking to kids,...so I guess its incidental. I jot things down on bits of paper and try to cover that whether its informally or as a class group. (Katy, Waimaka)

And Luke said

I've got an observation book.

[Mary] And you make notes in there?

Very occasionally. I know I should do it more, so that's half the battle [carrying out the assessment], the other half is actually getting it onto paper. (Luke, Waimaka)

Luke went on to say that he felt tests were also useful and “quite rewarding”.

### Three Approaches to Classroom Assessment

From the analysis in Chapter Six and outlined above, I theorised that these teachers seemed to be using at least three different approaches to their classroom assessment practice. I labelled these “unit” assessment, “head note” assessment, and “integrated systematic” assessment.

#### Unit Assessment

The teachers I characterised as using a unit assessment approach mainly described their assessment practices as occurring at the end of a unit of work and usually in terms of how well the children had met particular achievement objectives within that unit of work. Typically, these teachers planned from the New Zealand curriculum documents, choosing appropriate achievement objectives and sometimes reducing these to several specific learning outcomes from an appropriate level. Unit assessors mainly saw teaching and assessing as separate activities, describing instances of assessment as reasonably formal, planned events. For example, one teacher in this category told me

While we're all doing fairytales they'll be all meeting different needs within the fairytales unit. So I suppose I'll assess on that. Can this group do blends? Is this group able to use a contents page? Is this group able to tell me about the characters in the story? (Rose, Arawhata)

These teachers employed a coded scale to record each child's achievement of each objective (e.g. “has not yet achieved outcome”; “has met the teacher's expectations for this outcome”; “exceeded the teacher's expectation for this outcome”). These teachers also appeared to collect many samples of students' work as evidence of learning and as a record of test outcomes. They used this

information mostly for writing reports for parents or forming and reforming groups of students with similar teaching needs within the class.

I tentatively grouped Viv from Waimaka School and Rosemary, Rose and, possibly, Sam from Arawhata School as mainly using a unit assessment approach.

### Head Note Assessment

Teachers with a head note approach to assessment tended to rely on their memories of what children could do. Some also jotted notes during teaching. Although they referred to the achievement objectives in their planning, they kept meagre systematic records. In fact, some of these teachers stated that records interfered with teaching. They tended, in their own words, to focus more on “teaching” rather than “assessing” or recording. These teachers described assessment during teaching, focusing on individual learners and moving them along in different aspects of their learning as necessary

I work with the children on the floor doing work with them and we'll go over the work. It might be decompositions, subtractions, and I discuss and work with them until they can do it. Then they go away and do practice. I'll get their book and then afterwards find out how they have gone, if they understand it then I can move them on, if they don't I'll get them back and see them. So that's a way of finding out where they're at and that's on a daily basis and that's extremely effective because you cater for the children's needs. (Chris, Arawhata)

In this approach assessment was built into teaching but with little or no recording. These teachers generally kept samples of each student's work as a record and for reporting purposes. At least once a term or twice a year, these teachers set tests or assessment tasks and checked off progress against the achievement objectives. Luke and Katy at Waimaka School, and Chris at Arawhata School, gave responses that seemed to fit within this category.

### Integrated Systematic Assessment

These teachers generally spoke about using assessment information as they taught, while also acknowledging that they used it for reporting as well. They said that they recorded their assessment information systematically, typically against criteria or specific learning outcomes. They also described ways in which they used this assessment information to modify their ongoing teaching for individuals and groups. Sometimes this took place during teaching, but it was also often used for daily and weekly planning. For example, one of these teachers described how she achieved this almost individualised approach to assessing and teaching with her data book in hand.

So it's with me, nice and close to me whenever I want it...I know the document [data book] so well...that as the children are doing their work, I can quickly just note down in it what's happening or what they need. I have it on hand. I couldn't work without it. I need to refer back every so often. (Anne, Arawhata)

The data book this teacher referred to was a hard-backed spiral-bound book that she had prepared ahead of the teaching year. At the front she had written all the children's names down the outer edge of the left-hand page. She had cut away the outer edge from each of the remaining left-hand pages so that the list of names was in full view at whatever page the book was open. She had divided the book up into separate sections for each of the curriculum areas and essential skills, writing specific learning outcomes across the top of each page and leaving room for comments at the right-hand edge. Although this book was similar to the roll books referred to by other teachers, it was far more elaborate and contained specific learning outcomes, particularly in literacy and mathematics.

More than the other two groups, these teachers described collecting and using assessment information during teaching. Anne from Arawhata School, and Marvin and Rosie from Waimaka School, could be described as within this category. Lisa, the first-year teacher I had interviewed at Arawhata School, gave responses that could not be neatly classified within just one of these approaches. Her responses seemed to fit into at least the integrated systematic and the unit assessment categories. I therefore noted her approach as mixed.

## Discussion

The three approaches to classroom assessment proposed above had some similarities with the models of teacher assessment (TA) described by Gipps et al (1995). As explained at some length in Chapter Four, three models of teacher assessment practice were distinguished in the study: Evidence gatherers; intuitives; and systematic planners. The intuitives in Gipps et al. (1995) relied on their memories of what children were able to do and rejected systematic recording during teaching because they saw it as interfering with the teaching process in much the same way as the head note assessment group did in this New Zealand study. However, in these first two rounds of interviewing the teachers in the present study were not as articulate about defending a child-centred view of education as the children's needs ideologists of Gipps et al.. Like the tried and tested practitioners, the head note assessors tended to use teacher-made tests and worksheets to check up on progress at regular intervals but, in contrast to the tried and tested practitioners, said they mainly used assessment during teaching formatively.

The group of teachers identified as evidence gatherers (Gipps et al., 1995) had some similarities with the unit assessment approach outlined above. Like the evidence gatherers, the unit assessors did appear to believe that the purpose of assessment was mainly to check up on the objectives that had been taught in a specific unit, usually after the unit had taken place. In common with the evidence gatherers, the unit assessors tended to record progress against particular achievement objectives summatively, often at the end of a unit or at the end of the week or term. But, although the teachers at Arawhata and Waimaka Schools gathered evidence to support their summative decisions, they did not appear to go overboard in gathering as much evidence as they could, as the evidence gatherers were reported as doing (Gipps et al., 1995). In fact, the teachers classified mainly as unit assessors were just as keen as the other teachers at Arawhata and Waimaka schools to streamline their portfolio and profiling systems.

The third group identified by Gipps et al. were systematic planners. They were called this due to the fact that they planned for assessment on a systematic basis and that planning had become part of their practice (Gipps et al., 1995, p. 42). This group was subdivided into two further categories: “systematic assessors” and “systematic integrators”. The third approach identified at Arawhata and Waimaka Schools as an integrated systematic approach was also a more systematic one than the other two. Like the systematic planners (Gipps et al., 1995), teachers classified as using an integrated systematic approach to their assessment practice systematically planned for the assessment they intended to carry out. They, too, mainly focused on using assessment for formative purposes, recording the information for each individual student in a way that allowed the next teaching steps to be based upon it. Like the systematic assessors (Gipps et al., 1995), some of the Waikato teachers set aside specific parts of the day or week for assessment while, at other times, they built their assessment into the ongoing teaching programme in the way that systematic integrators did (Gipps et al., 1995). However, it was not possible to separate out two sub-groups of assessors and integrators as Gipps et al. (1995) did. All of the teachers that seemed to be using an integrated systematic approach incorporated aspects of systematic planning and assessment, and, at times, carried out their assessment practice during teaching

Pryor and Torrance (1996) reported that checklists, or, as they called them, ticksheets, had been a predominant feature of teachers’ assessment practice in the early years of the national curriculum in England and Wales, often promoted by education advisers and commercially produced by some schools. This checklist phenomenon was evident in the practice of most of the teachers at Arawhata and Waimaka Schools. However, as explained above, there were some significant differences in the form and use of checklists. Some seemed to be of the kind described by Pryor and Torrance (1996) as reducing the record-keeping process to a series of yes/no tasks, although in the Waikato schools a three-point scale seemed most popular. But others, principally those used to tune teachers in to important steps in the learning sequence, appeared to be of a similar kind to those Pryor and Torrance promoted for recording progress in reading. Not only did

these provide a record of progress, but sometimes a record of the interaction that had taken place between the pupil and the teacher, and the next teaching steps were indicated by the progressive nature of the more or less complete set of learning outcomes already entered in the book.

Due to these similarities and differences with the findings of Gipps et al. (1995) and Pryor and Torrance (1996), I decided to investigate the teachers' practice in more detail. I considered the methods used in previous research into teachers' assessment practice, including teachers' responses to quotations from other teachers and vignettes constructed from interview data (Gipps et al., 1995; 1996), interviews and observations of teachers in action in their classrooms (Bell & Cowie, 1997; Broadfoot & Pollard, 1998; Gipps et al., 1995; Torrance & Pryor, 1995), videos of teachers undertaking assessment in infant classrooms (Pryor & Torrance, 1996; Torrance & Pryor, 1995) and action research to explore the use of portfolios in teachers' own classrooms (Crumpler, 1996; Fenner, 1995).

I decided to pilot teachers' responses to teacher quotations as well as another set of interview questions to see which would provide the richer data to extend my understanding about the approaches I had identified. These two methods were chosen as practical strategies following discussions with authors from all of the research studies listed above except Fenner (1995). While videotapes of classroom action supplemented by interview and discussion with participants would most likely provide extremely detailed analysis, this method was extremely time consuming and beyond the resources at my disposal. Using an action research methodology was considered, but, as explained in Chapter Five, I believed my outsider positioning in the research process excluded this type of investigation. The teacher quotations activity (see Appendix E) and a further interview were piloted with a teacher from another Waikato school to see which might be most productive. While the teacher quotations activity did seem to indicate the approach that the teacher might be using, the interview questions indicated the approach and provided much richer information about other aspects, such as why the approach was used and how. Therefore, these interview questions formed the core of a third interview (see Appendix D). Seven teachers, drawn

from the original 12 interviewed at Arawhata and Waimaka Schools, were interviewed again in 1997 in order to amplify these findings. These interviews and the findings from them are described below.

## **Delving Further into the Assessment Approaches at Arawhata and Waimaka Schools**

Two teachers who characterised each of the three approaches outlined above were selected for a third interview. In addition, I decided to include the teacher who had seemed to use a mixture of all three approaches. I tried to choose one teacher from each category from each school, but due to staffing changes, both teachers in the unit assessment approach category were from Arawhata School. The teachers I interviewed for a third time in Term 2 1997 were as follows:

### **Unit Assessment Approach:**

Rosemary (Arawhata School)

Rose (Arawhata School)

### **Head Note Assessment Approach:**

Katy (Waimaka School)

Chris (Arawhata School)

### **Integrated Systematic Assessment**

Marvin (Waimaka School)

Anne (Arawhata School)

### **Mixed Approach**

Lisa (Arawhata School)



The information gained from the third interview confirmed that there seemed to be distinctive differences between the ways these teachers approached classroom assessment. Unfortunately, the interview with Chris failed to be recorded by the tape recorder and valuable information about Chris's approach to assessment was lost. I asked Chris if we might do the interview again but he had recently taken on a senior management role and was not keen. However, I had made notes during this interview that I was able to use to confirm my impressions about his approach. All the other interviews were successfully recorded, transcribed and analysed using NUD\*IST software.

The information from the third interview was coded in relation to:

- How the teachers reported using assessment for formative and summative purposes;
- Why, how, where and when the assessment information was recorded and how it was used;
- How the assessment information was interpreted;
- Whether, and if so, how the assessment information was monitored against the achievement objectives from the curriculum;
- How each teacher believed he/she learnt about assessment theory and practice.

In addition, the interviews produced information about:

- Repetitious surveillance of learning often during teaching by both teachers and peers;
- How assessment was timetabled into the daily programme;
- How teachers used assessment to scaffold learning;
- How teachers divided the curriculum content to ensure coverage.

The evidence from this third round of interviews confirmed the categorisation of the three approaches to assessment outlined earlier in this chapter. Lisa, who had been a first-year teacher at Arawhata School during the first and second interviews and whose responses to those interview questions showed a mixed response, demonstrated an integrated systematic approach to her practice in the third interview. In explanation, she had moved into a shared

teaching situation with Anne at the beginning of her second year of teaching. Under Anne's guidance, Lisa had quickly learnt how to implement an integrated and very systematic approach to her teaching and assessment, and her interview responses clearly reflected this approach.

Interestingly, however, the responses from the third round of interviews demonstrated that although the teachers had preferred ways of undertaking assessment, they varied their approach according to the context. For example, some used an integrated systematic approach for literacy and mathematics but tended to use more of a unit assessment approach in other curriculum areas. Others varied their approach according to the time of day or their confidence with the subject matter. In particular, although those teachers using a unit assessment approach described their practices as outlined above, they also explained that they had a progression in their minds, often different from the achievement objectives they were ticking off, which they used to discipline learning. In this way they appeared to have a parallel approach, formally monitoring the outcomes of each unit against their achievement objectives while informally responding to children's learning in response to a progression of learning they held mentally.

The teachers interviewed who appeared to be using a head note assessment approach also had something of a parallel approach. While not recording systematically or against learning outcomes or the national achievement objectives, Katy and Chris did make notes, list children with similar needs, and mark the children's work – strategies that they felt assisted them to remember what they needed to know to make decisions about future teaching strategies.

From this third round of interviewing it became clear that all three approaches allowed the teachers to meet the accountability demands of assessing individual student learning, monitoring against the achievement objectives, and recording and reporting progress. The differences between them related to the emphasis given to recording the assessment information they gathered. Specifically, these were differences in how systematically they had planned to collect this information, when they collected it, how and when they recorded it, how they used it to inform their teaching, and when, to whom and how often they

reported this information. An integrated systematic approach to assessment regulated the collection, use and dissemination of the information in a way that allowed these teachers to individualise their teaching to a greater extent than did the other two approaches. While, in contrast, unit assessment led to a more group-centred approach, it was difficult to make any generalisations about the effect a head note approach had on teaching and learning.

Although the three rounds of interviewing described to this point investigated the practice of these teachers in some depth and there were some obvious similarities with the findings of other international research (for example, Broadfoot & Pollard, 1998; Gipps et al., 1995; Pryor & Torrance, 1996), it was still difficult to know whether teachers at other schools in New Zealand were using similar approaches. I was keen to find out whether these assessment approaches and practices were common to teachers in other New Zealand schools. In 1997 an opportunity arose to investigate these issues with teachers at schools throughout the country. As part of the Recording and Reporting Trial (M. Hill, 1998a) I was contracted to carry out for the Ministry of Education, I was able to interview 20 teachers in the schools I visited to support and evaluate the trial. The results of these interviews are reported below.

### **Other Teachers, Other Schools**

The schools that took part in the Recording and Reporting Trial (M. Hill, 1998a), described briefly in Chapter Four, were, like the two Waikato case study schools and nearly every other school in New Zealand at that time, all in the process of developing their school-wide assessment, recording and reporting systems. The schools involved volunteered to trial either or both of the draft versions of the cumulative record cards and report forms provided by the Ministry of Education (see Appendix F for a fuller description of this project). These schools were situated throughout New Zealand, represented a range of deciles and sizes, and varied from rural through to urban in location. One aspect of the trial involved my travelling to many of these schools to assist the teachers to implement

the record cards and report forms, investigate how they were using them and explore other aspects related to recording and reporting assessment information. Twenty teachers from 13 of these schools agreed to take part in semi-structured interviews with me about their assessment practices. Table 7.1 below sets out details about the schools these teachers were from and the number of teachers interviewed at each school. The terms and codes used are explained in the footnotes to Table 7.1.

Table 7.2 summarises information about each of the teachers in the national sample described above. The teachers' names are pseudonyms to protect participants' privacy, as explained in Chapter Five. In most cases the teachers chose their own pseudonyms but some did not and, in one case, a name already used for a teacher in one of the case study schools was selected. I therefore selected pseudonyms when necessary.

**Table 7.1**

Overview of the schools from which the teachers were drawn in the national interview sample

| <b>Type of school</b>      | <b>Decile</b> | <b>Location</b>  | <b>Size</b>          | <b>No of tchrs interviewed</b> | <b>School code</b> |
|----------------------------|---------------|------------------|----------------------|--------------------------------|--------------------|
| Contributing <sup>15</sup> | 1             | South Auckland   | large <sup>16</sup>  | 1                              | NMC <sup>17</sup>  |
| Contributing               | 6             | South Auckland   | large                | 1                              | NFC                |
| Full primary <sup>18</sup> | 8             | Taranaki         | small <sup>19</sup>  | 2                              | NEP                |
| Contributing               | 6             | Taranaki         | medium               | 2                              | NWC                |
| Contributing               | 2             | Bay of Plenty    | large                | 1                              | NAC                |
| Contributing               | 5             | Bay of Plenty    | medium <sup>20</sup> | 1                              | NOC                |
| Intermediate <sup>21</sup> | 1             | South Auckland   | large                | 1                              | NWI                |
| Full primary               | 4             | North Otago      | medium               | 2                              | NBP                |
| Contributing               | 4             | Mid Canterbury   | medium               | 2                              | NTC                |
| Intermediate               | 4             | Mid Canterbury   | large                | 1                              | NTI                |
| Full Primary               | 10            | Central Otago    | large                | 2                              | NQP                |
| Full Primary               | 7             | Southland        | medium               | 2                              | NOP                |
| Contributing               | 8             | South Canterbury | small                | 2                              | NDC                |

---

<sup>15</sup> Contributing primary schools teach children from New Entrants/Year 1 (5-year-olds) to Year 6 (10/11-year-olds).

<sup>16</sup> More than 15 teachers.

<sup>17</sup> School code indicates N for national sample, then the first letter of the school name, followed by the type of school (I–intermediate, P–full primary, C–contributing)

<sup>18</sup> Full primary schools teach children from New Entrants/Year 1 to Year 8.

<sup>19</sup> Between 1 and 7 teachers

<sup>20</sup> Between 8 and 14 teachers.

<sup>21</sup> Intermediate schools teach children in Years 7 and 8 (11/13-year-olds).

**Table 7.2**

Overview of the teachers age, gender, years of experience and class taught in the national interview sample

| <b>Tchr<br/>Code</b> | <b>School<br/>Code</b> | <b>Age<br/>Group</b> | <b>Gender</b> | <b>Years<br/>as tchr</b> | <b>Class level<br/>taught</b> |
|----------------------|------------------------|----------------------|---------------|--------------------------|-------------------------------|
| Francie              | NMC                    | 40-50yrs             | F             | 20                       | Year 8                        |
| Dale                 | NFC                    | 30-40yrs             | F             | 14                       | Years 1 and 2                 |
| Rae                  | NEP                    | 40-50yrs             | F             | 20                       | NE-Year 2                     |
| Tom                  | NEP                    | 40-50yrs             | M             | 23                       | Years 3-5                     |
| Kate                 | NWC                    | 20-30yrs             | F             | 5                        | Years 4 and 5                 |
| Muff                 | NWC                    | 30-40yrs             | F             | 17                       | Years 1 and 2                 |
| Maria                | NAC                    | 30-40yrs             | F             | 12                       | Years 1 and 2                 |
| Alison               | NOC                    | 30-40yrs             | F             | 14                       | Year 3                        |
| Sharon               | NWI                    | 40-50yrs             | F             | 20                       | Year 8                        |
| Jessie               | NBP                    | 30-40yrs             | F             | 15                       | Year 5                        |
| Leila                | NBP                    | 40-50yrs             | F             | 17                       | Years 6 and 7                 |
| Pam                  | NTC                    | 40-50yrs             | F             | 18                       | NE and Year 1                 |
| Sally                | NTC                    | 20-30yrs             | F             | 2                        | Year 1                        |
| Andrea               | NTI                    | 20-30yrs             | F             | 1                        | Year 8                        |
| Delia                | NQP                    | 50-60yrs             | F             | 25                       | Years 2 and 3                 |
| Vanessa              | NQP                    | 20-30yrs             | F             | 8                        | Years 4-6                     |
| Bev                  | NOP                    | 30-40yrs             | F             | 18                       | NE and Year 1                 |
| Tammy                | NOP                    | 20-30yrs             | F             | 2                        | Years 3 and 4                 |
| Rachel               | NDC                    | 40-50yrs             | F             | 15                       | Years 1-3                     |
| Penny                | NDC                    | 30-40yrs             | F             | 15                       | Years 4-6                     |

## The Interview Questions

The interview questions used with the national sample of teachers were very similar to the questions used in the third round of interviewing at Arawhata and Waimaka Schools.

### 1. Ways of doing teacher assessment

Many teachers have told me that they use assessment in their classrooms to find out what the individual learning needs of their students are. Can you please describe for me how you find out what these individual needs are, for example, in written language or reading?

Is it the same for all curriculum areas? For example, what about mathematics? Science?

### 2. Ways of making decisions about children

How do you know what to do next, once you know what a child can or cannot do?

Can you tell me how you group children, if you do?

How do you know that/when you have sufficient evidence to say a student has achieved a particular objective?

### 3. Ways of recording information

Can you tell/show me how you record the assessment information?

Why do you record information in this way?

Can you show me how you record progress in terms of the national achievement objectives?

Do you use “levels” as described in each national curriculum document to record individual children’s progress through the curriculum?

Why/why not?

### 4. Ways of gathering evidence

How do you gather evidence about what children can do?

What evidence do you collect?

How do you store it?

What is the evidence used for?

### 5. Moderation

Do you have any ways of knowing if your decisions about student achievement are similar to those made by other teachers in your syndicate or school?

If so, what are they?

Why do you do that?

#### 6. School systems of assessment and record keeping

Can you describe any changes that have been made to your school's recording and reporting system since the new curriculum has been introduced?

Why have these changes occurred?

Do you intend making any improvements to your school recording and reporting systems in the near future?

If so, what changes and why?

#### 7. School evaluation assessment

How does your school monitor the progress of groups of students in order to evaluate the effectiveness of teaching and learning?

#### 8. How teachers learn about assessment

Can you tell how you learnt about the assessment practices you use?

From your perspective, what helps you most to learn about and improve your assessment practices?

Some teachers I have interviewed indicated to me that they believed the assessments they were expected to carry out filtered down from above, often originating with the ERO. Where do you think your assessment expectations come from?

#### 9. Self-assessment

Do you encourage students to use self-assessment?

If so, how?

Why/why not?

Running the interviews as conversations allowed me to probe into the approach each teacher described for assessment, recording and reporting. Each audiotape of every interview was transcribed and sent back to the teacher concerned for checking and confirmation. The NUD\*IST software programme (as



described in Chapter Five) was used to analyse the responses from these teachers and to identify how these teachers responded to the assessment discourses impacting on their practice.

### The Teachers' Responses

Using the NUD\*IST software programme, the responses to each question by the 20 teachers in the national sample were examined line by line and coded according to:

- Whether the practices used were mainly for formative or summative purposes;
- Whether assessment criteria were related to the achievement objectives, specific learning outcomes or graded judgements;
- What sort of evidence of learning was collected;
- If, how, why and when the information was recorded;
- How the assessment information was interpreted and used;
- How the teacher believed he or she had learnt about assessment;
- How surveillance of the teacher was implemented and who did this.

These were categories that had emerged from the earlier interviews at the two case study schools. They that allowed me to generate the propositions about the three approaches to assessment explained earlier in this chapter. But, in addition, in these interviews with teachers from a larger number of schools, there was information about some teachers' views about learning, their feelings about assessment, where they believed assessment expectations came from, moderation between teachers, school-wide assessment and changes that had occurred since 1993. These aspects were coded as well. That is, as information emerged on these topics from the data, new categories were formed to accommodate it.

### The Effects on Teachers' Assessment Practices

Some clear patterns emerged from the responses to the questions in this interview. The results indicated that the approaches to assessment identified in the case study schools as unit assessment, head note assessment and integrated systematic assessment could also be distinguished from the responses the wider sample gave. Each of these approaches is amplified below.

#### Unit Assessment

Nine of the 20 teachers appeared to be using an approach to assessment commensurate with that described above as a unit assessment approach. These teachers were from all levels of the school (Year 1 to Year 8). They mainly described their assessment practices as occurring at the end of a unit of work and recorded achievement in terms of how well the children had met particular achievement objectives set for that unit of work. Typically, these teachers planned from the New Zealand curriculum documents, choosing appropriate achievement objectives and sometimes reducing these to several specific learning outcomes from an appropriate level. For example

In oral language its got whether they're speaking clearly, whether they can sequence it, whether they seem to understand if its like current events. Also [I'm] doing a listening one on the people who aren't up the front. Are they listening, do they ask sensible questions, or did they not listen or ask something they've already been told? And I write little comments on the side. (Jessie, NBP)<sup>22</sup>

This occurred across the curriculum.

Like in maths, we have the set objectives we have to teach in all the different units.... These were my two objectives and then I'll just go through and tick them. (Andrea, NTI)

And these responses demonstrated that these ticking approaches were associated with the units that were planned.

---

<sup>22</sup> See explanation of school codes in Table 7.1.

I do use a unit plan. I write it with all my learning outcomes and achievement objectives plus daily lessons. The achievement objectives go in my roll book. I just use a pre-test as an idea of where they're at. I don't write that information down because I haven't actually taught them anything so it's just an idea for me of where they're at. I record at the end of a unit. After a post-test or an observation or whatever, after about two to three weeks and basically I use dot, tick, cross. (Tammy, NOP)

Like the teachers using a similar approach in the case study schools, the teachers using a unit assessment approach in the national sample mainly saw teaching and assessing as separate activities, describing instances of assessment as reasonably formal, planned events. Another similarity between the teachers in the case study schools and those in the national sample was that almost without exception, these teachers employed a three- or four-point scale to record each child's achievement of each objective.

Dot, slash and a cross. A dot is starting, a slash is developing and a cross is when they have mastered and are competent in what they are doing. (Sharon, NWI)

They usually had uniform school-wide documents for recording these checklists in. A term often used for the books containing these grids of specification was "roll book".

In my roll book I'll have the learning objectives or what I'm actually focusing on with a tick, pluses and minuses. I try to keep it to the yes, no, sometimes, always, that sort of thing. (Vanessa, NQP)

Some of the teachers had modified the three-point scale to include a more extreme category, as in this example.

A dot means they can't do it, a tick means it's good, a tick with a tail means very good, they're pretty good at it, and with a slash through it, they're really good at it. [For example] we did overarm throw, their style, distance, whether they could catch a ball that was coming down on top of them, whether they could catch a ball when it's coming across. (Jessie, NBP)

Several of these teachers indicated that they assessed at one level of the curriculum, or at the most, across two levels. To get around the difficulty of working at different levels of the curriculum several teachers described cross grouping across classes.

[Cross grouping] happens within our syndicate and the extension group is between two syndicates of Form 2 [Year 8]. So we have the four classes divided into five groups. That's maths assessment and their grouping is done on the basis of the checkpoint test that we do. And for English basically it is in three reading groups and three spelling groups that we get from the PATs. The PATs are done at the beginning of the year and from then on we see the children's work in class and how much of an improvement has been made. (Sharon, NWI)

As explained in Chapter Six, PATs are Progressive Achievement Tests. These are standardised norm-referenced achievement tests in reading vocabulary and comprehension, listening skill, mathematics and study skills, available to schools nationally but optional. Most have been recently revised to provide culturally appropriate material and up-to-date norms.

Teachers at other levels of the school also worked mainly at one level of the curriculum as well.

Basically we work at level 2. That's where I plan my lessons for but I have probably got children who are at the end of level 1 and who take up more time and I've got children who I do extend on to level 3 but that is extension activities.

[Mary] So, basically your assessment is recorded against level 2?

Yep! I mean I could put a level three assessment down there but that would mean extra objectives along there. For the likes of reading we work within a syndicate and the kids go back and forwards or whatever for certain subjects so they don't just have one level [for reading]. (Tammy, NOP)

In almost all the schools where unit assessment was taking place, the school policy indicated that recording information in a coded way against the achievement objectives was expected. This came through in statements like the following.

We have here the triangle assessment, which is a quick way. The dot is no attempt at all, then the start of the triangle means that they are attempting, and it works through until the filled in triangle is highly skilled at whatever they are attempting. It is really quick and we use that a lot. (Muff, NWC)

Included in these schools policies, whether written or simply expected, were samples of the children's work collected at regular intervals and classified against the outcomes listed with the relevant code.

We have set language samples which is school-wide. We get...three samples of their writing throughout the year and use a checklist. The key is a school-wide thing too, so dots [mean] commencing, tick [means] inconsistently and u [means] usually. Projects we use triangles. (Jessie, NBP)

Some schools had reclassified the curriculum achievement objectives into levels within levels. For example, this school had designed a complete assessment system with activities to use in order to classify each student within each level.

We've introduced the monitoring book. And we've identified the learning outcomes of the early, emergent and fluent at each of the different levels. Now we've formed a booklet of assessment activities for each of those levels, so at the end of each learning outcome, or at the end of each topic, they [the teachers] will give the students the assessment activity at that level. (Francie, NMC)

Many of these teachers also appeared to collect samples of students' work and test sheets as evidence of learning. They stored this evidence in individual student files, folders or scrap books and used it mostly for writing reports to parents or for forming and reforming groups of students with similar teaching needs within the class.

As discussed in Chapter Six, these coding systems used in unit assessment approaches seemed powerful in disciplining these teachers to categorize each of their students as performing at a particular level of the curriculum. As Luke (following Foucault) argued,

the effects and consequences of discourse are much more than just a naming of the social and natural worlds. Such disciplinary discourses do not yield social control in the coercive and

monolithic sense implied in Marxian analysis or even in some versions of classical sociology. It is not a “top-down” imposition of power. (Luke, 1992, pp. 6-7)

Practices such as the production of specified outcomes for each level of the curriculum, and even within each level, “make up practical ‘grids of specification’ for diagramming, classifying and categorising the subject in the social” (Luke, 1992, p. 7); thereby categorising not only each student, but also disciplining the teachers into practices which have the effect of colonising their practice. Several of the teachers using a unit assessment approach described to me how the school checked up to see that teachers were keeping their records according to the policy.

As the facilitator [of a curriculum area] I have a look and read the comments they are making. And at the end of the year I will collect them and keep them somewhere in case they may be required next year. (Francie, NMC)

As described in Chapter Six, most schools worked in syndicates. The teacher in charge of each syndicate was usually the one who checked up that the teachers were keeping their assessment records as the syndicate or the school as a whole had decided.

This is the way the school runs. I work in Betty’s syndicate. You’ve got to be understandable to her, but it’s also a personal thing too. To a certain extent you are influenced by the school, your unit, your syndicate leader and somewhere in there, there is still room for personal choice. (Tammy, NOP)

School-wide appraisal systems were also reported as a mechanism for making sure that each teacher was conforming to the assessment system within the school.

Well, we’ve got appraisals. I guess that’s when we would be going through and checking that there are not some teachers that are just leaving their children [behind]. Obviously every child is not going to improve in every unit but you would expect some improvement, so I guess it would come down to appraisal. (Andrea, NTI)

As a result of this categorisation and surveillance, these grids are used in schools in ways that generate self-surveillance. That is, the teacher internalises the disciplinary and cultural gaze as his or her own, taking on responsibility for monitoring his or her own morality, discourse and body (Luke, 1992).

While in the examples above these technologies of power were used to determine the conduct of individuals and submit them to keeping track of learners, mainly for accountability reasons, other teachers described working formatively with individuals within the approach. For example, while working within a school system that required school-wide unit forms to be filled in and checked off, one teacher stated

I had this argument with the ERO team because they were saying I needed to show exactly where the kids are on the national curriculum and it's pretty hard to.... I think you just look. It depends on what subject it is, like in reading there seems to be an obvious progression.

[Mary] And where does that come from?

I suppose just my own knowledge. Most of it would be prior knowledge. To be honest not much from the achievement objectives although you get ideas. In science and social studies I'm more aware of looking to see what we haven't done and fitting it in. A lot of it is just prior knowledge. I had a student teacher in my class for three weeks and she had real trouble with that. She would say "How do you know to ask them that?" or "What are you going to do tomorrow?" I guess as you teach you get more experienced with what's available. Some of it comes from resources but I think a lot of it is just your own knowledge. (Kate, NWC)

And another working within the same school explained how she sat alongside learners to move them forward in learning.

I try to do as much individualised reading with the children, just one-on-one first of all to really see what strategies they are using independently. And the same for writing. Just sitting with the children and seeing where they can go so that you're really seeing the process they are working through as they are doing it. To have a story handed in to you is one thing but to actually sit and see how they've moved through and if they have actually verbalised the word and listened to the sounds and if they can

articulate it slowly. You can't do that unless you're actually with the child, sitting with the group or whatever. So I try to do that either individually or just sit with a small group and observe there. (Muff, NWC)

This teacher went on to explain how she used her professional judgement.

When asked to explain what she meant by that she explained that

It's very much experience with a particular age group I think and I would say over the years my judgement and my expectations have changed a great deal and I have come to realise that the more I expect of junior children, the more they seem to rise to that particular level so my assessment level has become higher and higher for them as well and I think that's an experience thing really. I guess it's just being with children, talking to children, seeing how other people work and what their expectations are and I just think keeping on pushing the children really. Some of it's trial and error of course. Some of it you do make a mistake, you either push them a little bit too far or you teach them something they're not quite ready for. I think talking to the children is a good one. I've had reading recovery training and lots of things I probably took for granted before I did that training. Once I'd done the training, I'd notice that the children know their A, B, C's and the sounds that go with it and suddenly you'll have a child that will say "oh all those letters that are in that word are in the alphabet" and you think they hadn't really made that connection but that only really comes with talking to the child. Seeing how the child works, it's an individual thing. I don't think you'd come across that if you were just working in groups all the time. So I think the individual side of it's really important to be able to make some of those judgements and see where the children are heading. (Muff, NWC)

Other teachers noted how they also had a kind of progression in mind that they used during teaching, noticing and responding according to the context and the child concerned. When asked to explain how they knew what to do next when they had established what children could already do, several simply said "I just know." One teacher elaborated

Well, it's through reading. It's through having to work through all the curriculums at the same time. Having to read them all the time as references, having to plan from them, so you get to know what the material is and that's where you base what your children are doing. Other bits of reading that you pick up.



Professional reading, especially...some of the new handbooks. You pick up ideas and think that's a fabulous way of covering it and trial it that way. Sometimes it will be a complete flop and you won't do it again. Other times it will be a fabulous success and...I'll do it again. (Maria, NAC)

These teachers, in explaining how they “knew” learning progressions, how they watched children and noticed how they were progressing in particular aspects of learning, and how they used their “professional judgement” to intervene, demonstrated that while the pre-planned checklists and accountability mechanisms were pervasive they were not the only discourses of assessment impacting on their practice. Other professional discourses also influenced their practice. As well as their tacit craft knowledge (Bassey, 1999; House, 1998), acquired primarily through practical classroom experience, other assessment discourses, such as those explored in Chapter Six, were operating. For example, the teachers above explained how their professional reading, their knowledge of curriculum documents and subject matter, their professional training and advice from agencies such as the ERO all influenced their practice. It seemed, however, from the stories of these teachers at least, that unit assessors were heavily influenced by accountability discourses of assessment and were disciplined into recording systems that, in turn, had particular effects upon their assessment practices.

#### Head Note Assessment

Four teachers gave responses that indicated that they didn't often write assessment information down, preferring to hold most of their assessment information mentally, and testing or using check-up tasks towards the end of a period of time. One of these was teaching a New Entrant to Year 2 class. The other three were all teaching classes ranging from Year 3 to Year 7<sup>23</sup>. One explained how she couldn't hold everything in her mind and jotted numerous notes during teaching. She kept all these notes, occasionally referring to them if

---

<sup>23</sup> 7/11-year-olds.

necessary but not arranging them in any systematic way. Her folder bulged with small pieces of jotter paper bearing notes about children in every area of the curriculum.

Although referring to the achievement objectives in their planning, there were meagre systematic records. These four teachers seemed to believe that it was not manageable to write down all the assessment information they gathered and some stated that records interfered with teaching.

We were meant to put a code on it when we had done it to say achievement objective no. 4 – tick/dot/slash and date it. That would have taken me another two days for every entry but I didn't and I don't think anyone did, some may have but I don't think they did but instead I had it all to do at once so I didn't do that. (Leila, NBP)

While these teachers did explain that school-wide expectations for coding achievement of the curriculum objectives existed, and that they were expected to conform to these systems, they didn't participate. One teacher indicated clearly that for him, child-centred discourses of learning and assessment remained far more important than the requirement by both his school and the ERO to monitor each child at each level for each area of the curriculum.

I guess over the years we've developed here I guess its called like "child-centred" so basically when I try to organise a topic of work, I'm not that concerned about what the topic is but whether or not it provides opportunities for kids at different levels to have a go at all these different things. So [it's] very much integrated programmes with lots of things that kids at all levels can do. I guess you're trying to create an individual programme for each kid within your unit that you are doing.

[Mary] And would you monitor their progress in those things against the levels in the curriculum?

You're supposed to. We got an ERO visit at the beginning of this year and that's one of the things that came up in their report. They were not happy that we monitor sort of overall and didn't see us monitoring each kid at each level for each area. And when they asked me about it, I said "that's a really unrealistic expectation – that you can't possibly do it". I mean otherwise you would be assessing all the time. Every individual kid in every level in every area, but they were quite serious about that.

And I said to them “you show me a way of doing it and I’ll do it”. I’m quite happy to do it provided I find it a workable solution but it’s pretty difficult.

I had a new boy come into my room about the 2nd term who came over from Rotorua and the school that they were at they were doing this and he came with two clear files, chocker. I mean this was like the 1st term from that school and every second or third page was a different area with all the skills, all the guidelines, you know, right through every area for every kid. I mean I was thinking “my god I’ve got 34 kids, if I had 34 then I’d have 68 of those clear files to sit down and do, it just must...”

[Mary] It would take ages wouldn’t it?

Well that’s why a lot of those things in the end the teachers just go into automatic mode and go along and tick, tick, tick and at the end of the day they don’t really mean anything anyway.  
(Tom, NEP)

In contrast to his approach, this teacher believed that the assessment practices he saw other teachers using were driven by a managerial discourse of learning.

I find that a lot of the things that are being assessed at the moment are the things that are easily assessed and therefore they are sort of almost driving what’s being taught because you can only teach things that can be easily assessed because that’s what ERO seems to be wanting. They want to be able to see things that are easily assessed and then they can put all the things together and say “well there’s so many kids here that are not meeting expectations and what have you [done about it?]” But you do have to be careful that you don’t end up just teaching to those things because they are easily assessed and therefore you can teach them. (Tom, NEP)

And another teacher using a similar approach to assessment offered a similar explanation.

ERO have been strong on assessing to the achievement objective so we decided that’s what we would report on as well. There was a box that size for everything. It took me forever to start our reports because I thought I’m never going to get these done because I’m always running late anyway. When I sat down to do them, it didn’t take me as long to do them as I thought it would because I didn’t have to think about the whole job any more. I

just had to flick open a book and “what were they doing at speaking” and I wrote a specific thing about that. I just wrote to their achievement objective. I didn’t write a summary of their child’s speaking and I didn’t like it but that was what we had decided on. The same for algebra, I just wrote specifically about whether they could see sequences or patterns or whatever. Nowhere did I tell any of my parents “they enjoyed maths” or “tried hard”, “put in a big effort”, “they didn’t manage” and that annoyed me. (Leila, NBP)

Both of these teachers saw the ERO as driving school practice towards the checklisting-type approaches described above under the unit assessment approach. While the ERO criticise these forms of largely summative assessment as “quasi-mastery forms of assessment” that are imprecise when used to monitor the achievement objectives, and recommend that teachers use more formative strategies (Education Review Office, 1999), these teachers named the ERO as a driver of these quasi-mastery practices.

As with those who appeared to be using a head note approach in the case study schools described in Chapter Six, these teachers described using formative assessment during teaching, focusing on individual learners and moving them along in different aspects of their learning as necessary.

I guess it’s just built up over the years isn’t it, experience. I mean because you deal with kids all the time I suppose you meet such a broad range of kids that you get some expectation of what good kids are capable of and what poor kids need or where they are operating from so you make an instant assessment of where that kid is and you know that the next stage is such and such so that’s what you give them. If it’s too difficult you pull them back a bit and if it’s not difficult enough you move them on. I mean a typical example was this morning. We are doing a little reading lab at the moment just for interest sake. It’s a comprehension thing and it comes with an in-built test so the kids did that test two or three weeks ago and there were two kids today who brought up their results and showed me and I just looked at it and thought this is way too easy even though they came out on the test as to where they should be working. So I just immediately shoved them up ten cards and they went back and did it and their error rate I suppose improved which was good because it means that they’re starting to work. So instead of getting 100 per cent all the time, they were getting about 80-85 per cent. (Tom, NEP)

These teachers described developing professional judgement in much the same way as the teachers using a unit assessment approach had done.

You take your experience of different levels and different ages that you've taught and you build up a picture in your mind of where you think things are at. And you look at "OK, all these kids are 9 years old in my class" this is the range from here to here so maybe this is an average piece of work. I guess you make those judgements in your head all the time. (Penny, NDC)

The child-centred label was often applied by these teachers to their practice. For example, Tom used this term when he commented that over the years his school had developed a "child-centred" approach. But it was more difficult to understand how these teachers went about their assessment precisely because they had sparse documentation. To some degree this unwillingness to check up on individual achievement objectives could be interpreted as a struggle against the totalising force of the system. They were all teachers with many years of experience and did not appear to need a cover story for their lack of documentation. Clandinin and Connelly (1995) used the term "cover story" to mean the story used by a teacher, often when they were caught between competing stories (or, as I have termed it, discourses) in explaining their practice. In the context of beginner teaching and teacher education, they argue that there are competing stories of teacher education in the beginning teacher's professional landscape. Cover stories, they argue, are often used by teachers when they realise that what they do and believe are in conflict or competition with other practices that are expected by those above them in the education system. While at times they did seem to want to downplay this lack of records, one teacher told me

Yep and I'm about 4 months behind at the moment but don't tell Nick [the principal]. I'm usually about 6 months behind! (Tom, NEP)

These descriptions of the requirement to classify each child according to their level of the curriculum in every area as a new technology of assessment introduced following the marketisation of schools in concert with the new national

curriculum documents, exemplifies Foucault's description of discipline rather than overt power as a political strategy. He describes discipline as

...a multiplicity of often minor processes, of different origin and scattered location, which overlap, repeat, or imitate one another, support one another, distinguish themselves from one another according to their domain of application, converge and gradually produce the blueprint of a general method. (Foucault, 1977, p. 138)

Foucault goes on to explain that such techniques of power, in order to establish control,

were always meticulous, often minute, techniques, but they had their importance: because they defined a certain mode of detailed political investment of the body, a "new micro-physics" of power; and because, since the seventeenth century, they had constantly reached out to ever broader domains, as if they intended to cover the entire social body.... Describing them will require great attention to detail: beneath every set of figures, we must seek not meaning, but a precaution; we must situate them not only in the inextricability of a functioning, but in the coherence of a tactic.... Discipline is a political anatomy of detail. (Foucault, 1977, p. 139)

Tom recognised that by assessing each child against outcomes at each level of the curriculum in every curriculum area he would have no time left for using the child-centred practices he espoused. In this sense he exemplified the tension between the formative and summative roles (Black, 1998). In order to "hold the boundary between the two" (Black, 1998, p. 120) it was necessary for Tom to argue with the ERO and keep the fact that he was behind in his recording from his principal. Leila, too, explained her change to a less descriptive approach to recording learning information about children under pressure to conform to the accountability pressures in her school. In the attempt to produce standards, and to classify by children by those standards, managerial discourses of accountability, through the requirement to monitor every child against the stated achievement objectives, were in competition with the assessment discourses attended to by these head-noting teachers. They described having their own set of expectations for children in their heads, built up over years of experience, which they used

continuously to make teaching decisions during teaching. These descriptions included child-centred discourses and references to their own tacit craft knowledge. To deal with the requirements to monitor and record against the achievement objectives they either resisted recording in the required manner (Tom, for example) or complied superficially (as in Leila's case) to fill in information against "their" achievement objectives.

However, due to the lack of documentation, it was difficult for me to describe the assessment practice of these teachers in detail. By keeping assessment information in their heads, they resisted attempts to make their practice visible. Unless the ERO reviewers (or myself, for that matter) entered these classrooms to observe teachers' practice over a reasonable period of time, they could not report on the effectiveness of this teaching. This was in contrast to the practices of the other two approaches I had identified. This approach, it seemed, would also make communication and collaboration with other professionals more difficult.

### Integrated Systematic Assessment

The third approach identified earlier at the two case study schools I labelled an "integrated systematic" approach to assessment. Three teachers in the national sample, all teaching in the junior area of the school, described similar strategies and techniques. Two were experienced teachers in New Entrant/Year 1 classrooms and both were the assistant principal at their respective schools. The third was a young woman in her second year of teaching. She was teaching a Year 1/2 class and working closely with her assistant principal, one of the teachers also described as using an integrated systematic approach above.

In contrast to the head note approach described above, these teachers had a mass of documentation relating to ongoing progress, particularly through the early stages of reading, writing and mathematics. These teachers generally spoke about using assessment information for teaching purposes, while also acknowledging that they used it for reporting as well. They said that they recorded their

assessment information systematically, typically against particular aspects of learning.

I was looking at each child – that's the master coversheet we went through – and I do a check on how many high frequency words they might have, some are quite good. Could they write their name? Could they draw themselves? Running records we're mainly looking at there. Is directionality established? Is one-to-one matching established? Are they using picture cues? You're not really picking up their word knowledge at that stage but seeing what their book skills were. I did a letter identification test as in what's used on the observation survey. (Bev, NOP)

Rather than use the achievement objectives as their points of reference for learning, these teachers had developed their own set of learning outcomes. They did not keep these in their heads, as the teachers described above did, and could describe exactly what they checked up on and recorded.

We do an alphabet check to see if they know the letters and if they can also associate the letters with a key word. We also do a check on what basic words they can recognise. We have a reading graph which we fill in at the end of each term to say which levels they completed. (Pam, NTC)

These teachers typified the descriptions of progressive pedagogy provided by Walkerdine (1992). Commenting on the introduction of child-centred discourses of pedagogy, which was co-terminous with, and related to, the entrance of women into elementary school teaching, Walkerdine stated that

women could watch, monitor and map the child's development. Clipboard in hand, these scientific educators could survey each of their small charges, whose development was entrusted to their love.... The dream of ensuring each child's pathway to reason turned to the schoolroom, where pupils recited their lessons and moved up the form, into the classroom, a place in which each child was considered separately. Discipline became not overt disciplining but covert watching. Regurgitated facts became acquired concepts. Knowledge became naturalized as structure or process.... The classroom became the facilitating space for each individual, under the watchful and total gaze of the teacher, who was held responsible for the development of each individual. This assumed a total gaze, which could be stated, as



one teacher put it, as “knowing each child as an individual”. (Walkerdine, 1992, pp. 19-20)

Although Walkerdine considered that “it was always an impossible fiction” (1992, p. 19), her description holds remarkable similarities with the descriptions of teachers in this study using an integrated systematic approach.

A lot of it is observation and you see how they’re going and you’re just watching them. It’s kind of anecdotal and you’re thinking “gee I might need to do some running records on that child because I noticed they’re falling a bit behind in their group” or “they don’t seem to have come up with this”. (Sally, NTC)

These teachers described ways in which they used this assessment information to modify their ongoing teaching for individuals and groups. As with the teachers in the case study schools, sometimes this took place during teaching and it was used for daily and weekly planning.

For the week I was looking at capital letters and full stops. The next week, when I pick up their writing book if someone isn’t using it –its OK to forget maybe once or twice– but if they’re not using it all the time then I would notice that they need to do it again or I need to work with them again.

[Mary] So would you be recording that somewhere as you go along or not?

I usually just write their name beside my planning with a little asterisk like “look at so and so”.

In maths I have a thing that I use, we call it our log book. I write the objectives up the top or whatever I’m looking at for the week and then I write beside it. I notice whether the children are doing it and I’ve got space where I can write “so and so seems to be having difficulty with counting” and then the next day I think “oh I better check to see what they’re like”. (Sally, NTC)

As these extracts indicate, these teachers were building their daily and weekly teaching around what they noticed and recorded, either formally or informally, as teaching took place. One teacher also explained how she was able to keep track of rapid development for each child. Talking about keeping a record

of word recognition, vital for knowing when to begin early readers on structured reading material, she told me

This [record] is just for myself and I don't expect any other teacher to keep this but I need to keep a check on what high-frequency words they know because I use this in my planning. The kids can flick down the laminated card with high-frequency words on it pretty quickly. And they feel good "ooh I know that" and it often gives them a bit of confidence. Ideally I'd like to separate them [each child's sheet] out to separate piles but I find it's just in the interests of time, I haven't got it. I make sure every time I test I use a different coloured pen and then I know that on the 24/3 this child knew 21 sight words. So I can see what the rate of progress is. Once they hit around early to mid 20s then it's onto red books and then I don't continue to record those ones there that I know. (Bev, NOP)

From this explanation the purpose of keeping the records relates almost entirely to making decisions about teaching and for the teacher to keep track of individual students' progress. These teachers barely referred to using these detailed records for accountability purposes. Bev, in the statement above, however, did refer to the children gaining confidence from seeing the progress they had made. She had an extensive recording book that she kept her information in. Although this book was similar to the roll books referred to by other teachers, it was far more elaborate and contained specific outcomes, particularly in literacy and maths. Like Anne's book at Arawhata School, this one contained detailed information about progress for each child, mainly in literacy and numeracy.

Usually I'll comment on their cue usage, whether they're cross-checking in terms of visual information, what part are they focusing on. All parts of the word, one part of the word, are their attempts making sense. I'll comment on fluency, I'll comment on, for the little ones, direction, one-to-one matching, return sweep. The comments I will make will reflect the nature of what I've been trying to do. It is a lot of work but I do believe if anyone comes in and asks me about a child I can tell them boom, boom, boom and I can show them too, I've got the evidence there. So that's what I do for reading. For maths I'm going to have to actually extend on a page there. I won't get all the number 1s in. I break down each objective. There's a Wellington maths scheme where it actually breaks it down into steps which I found is really ideal because it's my first year of

teaching New Entrant maths. I was teaching further up the school so I'm having to break it down much much more. I break it down there. I plan for those. I usually plan for about 4 different groups and then I can usually see from about the first day the general activity "that kid will be in that group, that kid will be in that group, that kid will be in that group" and I tend to find that they don't change a lot. There might be some differences with some children. Some children are really good at number but not good at measurement but generally I find if they're pretty good at one maths strand. I double check. (Bev, NOP)

Bev understood the need to be accountable, but felt that if anyone came in she could demonstrate how she was monitoring and recording progress. Confident in her practice, she did not feel the tensions between formative and summative assessment in the way teachers using a unit assessment or head note approach had described. When necessary, she had adopted schemes of work from outside sources (e.g. the Wellington maths scheme) and adapted them to her purposes. In order to make her practice manageable, she divided the children into groups when necessary.

Another told me how she used her recording book in the covert fashion Walkerdine (1992) explained.

I have it with me as I'm walking around. It just sits beside me when I'm taking the group lesson but then when the children are doing their independent stuff then I walk around with it. I'm not sitting there checking them but I walk around and I think "oh so and so seems to be having a little bit of difficulty" and I just quickly scribble it in the book but I don't walk around with the book, looking in the book. Sometimes I just write it on my hand and think "gee I better do that again later" because I don't want to interrupt what the kids are doing because they're quite nosy and they often wonder what you're doing. (Sally, NTC)

It was interesting that, like the teachers at the case study schools, these three teachers tended to use a less detailed system for other areas of the curriculum. For example, one told me in

our unit studies, we were involved in doing achievement-based assessment. Basically we record a knowledge and skill objective and we break it down. I give my crew a fair bit of flexibility,

some people will use the triangle which is the same thing. (Bev, NOP)

Here the school system of checklisting against the achievement objectives took over from the teacher's own internal "progress map" (Masters & Forster, 1996). The "triangle" referred to was a three-point system, similar to the tick, dot, cross systems explained earlier. From the interview transcripts it was not easy to tell why the teachers reverted to using the achievement objectives, but it could have been that these were lower-priority subjects for these teachers, they had less knowledge of them, or that because it was impossible to individualise teaching completely, they restricted their intensive surveillance to that which was manageable. In Walkerdine's words, "knowing each child" was "an impossible fiction" (1992, p. 20).

#### Other Approaches

There were several teachers who appeared to use a mixed approach. For example two teachers described school policies that included checklists with codes for recording progress against the achievement objectives but explained their personal approach as kept in their head or in jotted notes:

Sometimes I'll do a pre-test and get some knowledge about that and we'll do a post one, get them to draw a picture at the beginning and the end, that's an easy one. We talked a bit about what they knew about space, worked on their ideas, what they thought they were and then compared it at the end. But basically, a lot of the things like reading and maths and writing are just individual and its day by day, you look at where they're at and then help them along individually so this [checklist] isn't a helpful... This [checklist] is all right for recording the big topics and subjects and things like that, that's easier to do but where it's little steps in reading and maths.... Just observing them while they're working and you just think "oops, that one doesn't know that, let's work on that one" and let the others go. If you're teaching in a group and one person doesn't know you keep them back a wee bit longer and we could go over it again and the others go off to an activity.

[Mary] Do you record any of that sort of stuff?

No. I'm not good at doing things like that. It's just in my head but then I remember it perhaps for the planning for next week so that as you plan during the week, you remember to do this and they need a bit more time on that. (Rachel, NDC)

It was also difficult to discriminate between approaches where teachers described a unit assessment approach within the school but appeared to be using the information they gathered about individuals and groups formatively. For example,

We do a lot of checklisting, well not a lot, but we do a fair wack of checklisting. We also have our own assessment criteria in things like handwriting, spelling that we are monitoring ongoing. Mathematics is the same. We have our criteria established. In written language we have learning outcomes that we are continuously assessing. These children, their needs change quite dramatically and so I found that just the recording of their names and their learning outcomes didn't give me enough information to follow that child through. I just found that that was a more concise way. I've always got that folder with me, and when I'm taking that group I'm aware of what I'm looking for and I also keep a few anecdotal records in there with those children as I see those things happening.

Let's say I have a little boy who's having trouble with his alphabet. What I've done there is just noting down the new alphabet letters that he's achieving then I can group together particular sounds that I can see that he's making a good attempt at, and ones I think he needs extra work on. That's just a very quick note down in the folder or because otherwise I'd be spending all my time writing and not watching what these kids are doing. (Dale, NFC)

Interestingly, as with the integrated systematic teachers, both of these teachers also taught junior classes. It seemed as though learning was especially rapid and observable in the early years of schooling and that these teachers needed to track learning individually to match teaching to it. Further up the school the teachers described only head note and unit assessment techniques.

## Summary and Comment

As this chapter has demonstrated, there appeared to be a large degree of resonance between the assessment practices of the teachers in the two Waikato case study schools and those of the 20 teachers in the national sample. Like the teachers at Arawhata and Waimaka Schools, the teachers interviewed throughout New Zealand seemed to be responding to a variety of assessment discourses. In particular, all seemed to be responding to the accountability demands of self-managing schools to some degree. Each of the schools had developed policies that ensured the teachers monitored individual student progress against the achievement objectives. Most recorded this progress against the curriculum levels, often using a three- or four-point code. Throughout the country this monitoring ensured that the teachers in some way assessed each individual child, compared his or her performance against a hierarchical scale of objectives, and recorded achievement accordingly. In doing so, the teachers used the achievement objectives from the curriculum or their own knowledge about children's learning to design "grids of specification" (Foucault, 1981) in order to classify learning and, consequently, the child. The progressions expected by the curriculum, as well as those internalised by the teachers, allowed the teachers to make normalised judgements about each of their students. Foucault (1977), in explicating how "discipline makes individuals" (p. 170) through hierarchical observation, normalising judgement and the examination, described how surveillance is critical to efficiency in teaching (p. 176). For me, the descriptions given by the teachers of the ways in which they carried out, recorded and acted upon assessment in their classrooms illustrated how such a relation of surveillance was inscribed at the very heart of their practices.

This surveillance of individuals had both productive and repressive consequences. The teachers in both the case study schools and the national sample described as using an integrated systematic approach, explained how internalising the "progressions of learning" assisted them to make judgements about the achievements of children in their classes and, in turn, make decisions about teaching in light of this knowledge. In this sense, the close supervision of each

child's learning was productive in that learning could be supported and enhanced almost continuously. The timetables, recording schedules, planning sheets and checking-up they had designed allowed them to run individualised and group programmes that integrated the assessment process within the teaching and learning programme.

Those teachers described as having a head note approach were also using their assessment information in teaching but seemed to be less systematic in their approach, making mental notes or jottings and recording systematically only occasionally, mainly for summative purposes. It was difficult to tell how their more intuitive approach guided and directed their teaching. However, as argued earlier in this chapter, these teachers were outspoken about the fact that they believed school and national policies of accountability appeared to be in tension with their teaching role. Understanding that using the achievement objectives, or learning outcomes derived from them, for the grids of specification (Foucault, 1981) often conflicted with the maps of progress these teachers held in their heads assisted me to understand why they were so critical of new assessment demands. The clash of discourses was thus inserted at the very heart of their assessment practices.

Those who explained an approach as that described as a unit assessment approach did not appear, at first, to be using formative assessment to the same extent as either of the other two groups. But, on closer examination, the fact that they described ways in which they had internalised learning progressions and that they used this "mental map" to guide their teaching suggested that they had something of a parallel approach. Although the use of end-of-unit checklists might seem to make the assessment process more manageable, the reality appeared to be that, because these were mainly summative devices, these teachers needed more intuitive formative strategies to support students' learning during teaching, adding to the complexity of the assessment practice. In other words, it seemed that the accountability discourses that produced school-wide policies of objectives to be covered and assessed with checklists at the end of units focused the teachers'

gaze on specific outcomes for the purpose of recording and reporting at the expense of meeting individual learning needs.

Some teachers described not only how they assessed children and acted upon this information but also how they disciplined the children they taught to use self-assessment: to keep themselves under surveillance and monitor their own performance against the learning outcomes. Drawing again on the “toolkit”, Foucault considered that the dividing practices through which the subject is constituted (in this case, the categories making up the monitoring sheets and checklists) were merely one aspect of the art of governing people in our society, “which needed to be complemented by an equivalent consideration of the ‘techniques of self’” (Smart, 1985, p. 108).

The concept of “techniques of self” refers to the means by which individuals can affect their own bodies, souls, thoughts and conduct so as to form and transform themselves. (Smart, 1985, p. 108)

Self-assessment, where teachers provide the criteria by which learners can judge themselves, compare themselves to a norm and transform themselves to meet it better, can be seen as an example of a “technique of self”. In fact, at its heart, formative assessment is a technique of self in that, as explained in Chapter Four

for students to be able to improve, they must develop the capacity to monitor the quality of their own work during actual production. This in turn requires that students possess an appreciation of what high-quality work is, that they have the evaluative skill necessary for them to compare with some objectivity the quality of what they are producing in relation to the higher standard, and that they develop a store of tactics or moves which can be drawn upon to modify their own work. (Sadler, 1989, p. 119)

From the responses of both the teachers at the case study schools and the teachers in the national sample, it seemed that they were positioned at the intersection of multiple and competing discourses of assessment. In order to become accountable, the self-managing schools these teachers worked in had taken the notion of the classification of learning into achievement objectives and



used these objectives, set out in increasing levels of difficulty, as a way of classifying and defining learners. As outlined in Chapter Two, this was not a completely new way of accounting for learning. Keeping School Records had been developed by the Department of Education in co-operation with educators as a way of defining learning and progress, but the learning goals (as the outcomes were termed) were only specified for primary schools in very general terms being couched at junior, middle and senior levels. They were listed for each of language, mathematics, science, social studies, physical education, health, art and music under three headings: knowledge, skills and attitudes. However, although schools were inspected prior to 1989, they were not reviewed on their effectiveness in terms of the learning goals listed in Keeping School Records.

Although teachers in primary schools prior to restructuring in 1989 had implemented progressive pedagogies and developed criterion-referenced systems of assessment (see Chapter Two), none were as systematic as the “grids of specification” that schools developed to demonstrate to the ERO and themselves that they were “monitoring student progress against the national achievement objectives” (Ministry of Education, 1993b, p. 4). And, as I have demonstrated above in Chapters Six and Seven, the collection-type curriculum (Bernstein, 1971), introduced after schools were made self-managing, in which

social order arises out of the hierarchical nature of the authority relationships, out of the systematic ordering of the differentiated knowledge in time and space, out of an explicit, usually predictable, examining procedure (Bernstein, 1971, p. 63),

has been productive of checklists which discipline teachers into checking off learning. The categories of approaches to assessment that I proposed in Chapter Six certainly resonated with those reported by the teachers in the national sample.

In order to confirm these propositions about the effects the various discourses of assessment were having on teachers’ practice I decided that I needed to observe teachers in action in the classroom. Although the descriptions teachers gave me in the interviews suggested distinct approaches to assessment, there were also comments within the responses that pointed to an even more complex picture in practice. To investigate in more depth how teachers were responding to the

multiple and competing discourses of assessment, I approached three teachers from the original case study schools and asked if I could undertake a week-long observation in each of their classrooms. In order to investigate possible different approaches to assessment, the three teachers I asked to take part in the participant observation part of the investigation were chosen because each seemed to represent one of the three approaches identified above. Chapter Eight describes how I went about this part of the investigation and reports the results of these week-long observations.

## **CHAPTER EIGHT**

### **ASSESSMENT IN CONTEXT: THREE PORTRAITS**

The case is no longer, as in casuistry or jurisprudence, a set of circumstances defining an act and capable of modifying the application of a rule; it is the individual as he may be described, judged, measured, compared with others, in his very individuality; and it is also the individual who has to be trained or corrected, classified, normalized, excluded, etc. (Foucault, 1977, p. 191)

It is necessary to restore the human subject at the centre – the suffering, afflicted, fighting human subject – we must deepen a case history to a narrative or tale; only then do we have a “who” as well as a “what” – a real person. (Sacks in Scott, 1999, p. 75)

In contrast to defining discourse as classroom talk, Foucault explained that discourses are tactical elements or blocks operating in the field of force relations; there can exist different and even contradictory discourses within the same strategy; they can, on the contrary, circulate without changing their form from one strategy to another, opposing strategy. (Foucault, 1978, pp. 101-2)

My intention was to restore the human subject at the centre (Scott, 1999) of the competing assessment discourses identified from the field work in this investigation. I set out to do this by investigating both rule following and decision making (Hammersley, 1981, cited in Torrance & Pryor, 1995), by examining practices in the classroom context for a sustained period, and by engaging teachers in conversation about that daily classroom action. From the interviews reported in Chapters Six and Seven it appeared that teachers’ assessment practices in self-managing schools were both constituted by, and constitutive of, multiple competing discourses of assessment. The assessment approaches described by teachers in these interviews demonstrated that they had found ways to meet the accountability demands to report achievement and progress as required by the school policies, appraisals and performance management systems.

In both case study schools, the teachers I interviewed saw assessment as tracking student achievement of the achievement objectives in the curriculum. But they also explained a major purpose of assessment as finding out where students were “at” so that, as teachers, they might take them on from there. This suggested that most of these teachers held constructivist views about learning and teaching. They believed that it was important to know what children’s current views and abilities were in order to take these into account in their teaching. However, how teachers then used this information for adapting their teaching practice was not really visible from the interview data. In order to understand what any particular strategy was, what its function was, it seemed essential to pursue an investigation of the patterns of intention and motivation that produced it. And this required the use of a wider range of methods than just interview and documentary analysis.

The idea that teachers were each at the centre of multiple and competing discourses of assessment, and the indications from the interviews that teachers responded to these discourses in relatively distinctive ways, led to the investigation of the practices of three teachers in more depth. These teachers were selected from the original participants at Arawhata and Waimaka Schools: Rose, at Arawhata School, had consistently explained her approach to assessment as that described above as “unit” assessment; Anne, also at Arawhata School, described her assessment approach as that labelled an “integrated systematic” one; while Luke, who had been at Waimaka School for the first two rounds of interviewing, seemed consistent in his description of that which I labelled in Chapter Seven a “head note” approach.

Each of these teachers agreed to be observed in action for a week. As explained in Chapter Five, a peripheral stance (Adler & Adler, 1994) was chosen for the observation. By “peripheral” I mean that although I had a close interaction with the teachers and their class members, as the researcher I was not involved in the core activities of the group. I shadowed each teacher during her or his working day, keeping a running record of the events of the day and noting any activity that appeared to me to be related to assessment. The teachers wore a small dictaphone tape recorder in a “bum-bag” round their waist with a condenser microphone

clipped to their lapels. This microphone was powerful enough to pick up all their talk as well as most of the conversations they were engaged in. Each tape was numbered and the number, side and time recorded in my running record of field notes as recording began.

I used the taped records, the field notes, examples of assessment interaction, assessment records and extracts from my interviews and my subsequent conversations with these teachers to examine their assessment action in practice in some depth in the following portraits. In constructing these stories of Rose, Luke and Anne, I have also interwoven interpretation and discussion of the discourses that appeared to be influencing their practice. I acknowledge that using this type of narrative approach has the potential to “romanticise the individuals and thus reify notions of a unitary subject/hero” (Munro, 1998). My intention is not to suggest that it is possible to capture experience and portray it as unitary, essential and universal, but rather to attempt to explain in more detail through these portraits how multiple, competing discourses impact on teachers’ practice, sometimes in unexpected and unintended ways.

Using case studies in this way can take us to places where most of us would not have the opportunity to go.... [They] allow us to look at the world through the researcher’s eyes, and, in the process, to see things we might not otherwise have seen.... And a third reason that the vicarious experience provided by case studies might be preferable to direct experience [is that] vicarious experience is less likely to produce defensiveness and resistance to learning. (Donmoyer, 1990, p. 196)

The assessment approaches I described in Chapter Seven and used to guide me in the selection of these three teachers were devised for their utilitarian application rather than their essentialism. As explained in Chapter Seven, these approaches were not completely distinct. Every teacher used aspects of the different approaches at varying times of the day, or varied his or her approach according to the curriculum area he or she was engaged in teaching. These three teachers, however, each described a more or less consistent way of approaching assessment in their classrooms. The close-up classroom observation of these three teachers enabled deeper probing into their classroom assessment practices as well

as providing evidence of how the approaches described in Chapter Seven were carried out in action.

In constructing these portraits, I was uneasy about speaking for the participants rather than have them speak for themselves. In one sense, constructing these narratives was attractive because it sought to give voice to teachers whom, to an extent due to the education changes of the 1990s, had been marginalised. But, in another way, the teachers might seem to be undermined by my writing “giving voice” to their practice. Giving voice by writing teachers’ stories as narrative was unsettling because it underscored my perception of those I was researching as not able to speak for themselves and conflicted with my beliefs about teachers as meaning makers, constitutive of their own discourses of assessment.

Although unable to resolve these tensions and contradictions, I proceeded, knowing that within each school context these teachers did have voice and that they had gained more understanding of the assessment discourses during the process of the research, enabling them to speak with more authority and to find spaces for change. In these co-constructed portraits, therefore, I have attempted to tell each story as I saw it, albeit by checking with the participants on many occasions to improve the accuracy of my account. Despite the reservations I had about the narrative process, I remained committed to putting the people back into this research report in order to restore the human subject attempting to juggle the demands of teaching while responding to the assessment discourses impacting on daily life.

## **Rose**

Rose had recently returned to teaching having spent the last few years raising her two boys, who were now students at Arawhata School. In total she had been teaching for around nine years and had a Year 2/3 class (mainly 6- and 7-year-olds) in 1996 when the first interview took place. She had a Diploma in Teaching, the basic New Zealand primary teacher qualification. She felt

moderately confident about assessment in the teaching process at the time of the first interview and believed that she had room to move in making decisions about the assessment practices she used in her classroom.

My interviews with Rose took place in a small cramped office cum resource room in a corner of the administration block at Arawhata School. From these conversations about her assessment practice, I had decided that Rose was using a unit assessment approach to her assessment; that is, she seemed to plan assessment objectives within the units of work she was teaching and check off how well each child was achieving these, often towards the end of a unit.

### A Checking Approach to Assessment

I have a little focus that I'm working on, whether it is capital letters this week or using more interesting words or dictionary skills. I do the focus on a Monday and then I work with the groups through the week and that'll be our focus. And then I'll collect the data on that, whether that's been successful. Whether that skill has been comprehended.... We have a sort of "no they didn't get it" which is a subtraction sign; a cross if they seem to have a strength in it. I tend to go for children who've been able to talk about it and show...really clear understanding. The others I tend to tick because I think, well, we've covered that topic, they've all talked about it and you hope that, yes, they've understood.... Sometimes I think we're gleefully ticking but do we really know?

Although Rose reported using this sort of checking approach she was not wholeheartedly in favour of it by any means. Rose seemed to be continually contesting the requirements to record assessment information. Although she said she believed the most valuable assessment to be that which tells where the children "are at and where they are going", linked to her planning about where she was "going to go from there", Rose described her practice as split between collecting and recording evidence and keeping much of the information about individual learners in her head, using it as she worked "alongside the children". The need to record assessment information was to a great degree a frustration that Rose struggled with on a daily basis.

Because assessment is only for that day, isn't it? Like you do a big push to see if they can use capital letters and you tick it all off. Have a look at their work a month later and I know [that] someone or other has lost that skill. But you've ticked on that day and they could do that.... In my roll book I think, well, I won't write it all out again, all those simple skills. I just go back and re-tick whether they still have that or whether they need to re-visit that skill.

### A Clash of Discourses

This struggle seemed to be a manifestation of the discourses competing for Rose's attention. On the one hand, her view of how learning occurred seemed to come from a constructivist discourse. She said that she wanted to know where the children were at in order to help them enhance that learning. But, at the same time, Rose was well aware of the requirement to monitor learning against the achievement objectives for policy reasons. Because of these expectations she had agreed with her colleagues to use a code to check learning off under the actual objectives straight from the curriculum documents. Rose explained to me that this was not working well as each objective was too general and covered a range of possible learning outcomes, but she felt bound to continue with this approach as it was her syndicate's decision. Nias (1989) found that when teachers re-entered the profession after a break, as Rose had done, they often had doubts about their professional expertise due to perceived changes in the system. She found that to meet the demands of the job they worked very hard and turned to their colleagues for assistance to know how (in their particular school) successful teachers behaved (Nias, 1989). Experienced colleagues in Rose's syndicate had decided to use the achievement objectives in this way, and Rose felt that she should identify with them and try this approach too.

But the progress map/time line she described in her head was not made up of a sequence of achievement objectives. One interpretation is that, by keeping her real assessment criteria covert (in her head), and using the achievement objectives to record against, Rose could not come in for criticism from her superiors or colleagues for not monitoring against the objectives as the NEGs and



the NAGs prescribed. As explained in Chapter Seven, Clandinin and Connelly's (1995) term "cover story" seems appropriate here, with Rose recording results against the achievement objectives while, in her head, living out a secret story by using her "map and time line" (Clandinin & Connelly, 1995). The tensions for Rose were not just between the summative and formative uses of assessment, although these were present, but also between what she believed was appropriate, what she felt she was expected to do, what was manageable and what actually worked.

### The Need to Account: "I'd have a printer in my head"

When asked what assessment practices she would implement in her ideal school, Rose told me that she'd have a printer in her head.

It would print out everything I'm thinking. I just think that we're going around all day observing little things and even though people say have little notebooks and take field notes, it doesn't work because it's...it doesn't work.

This desire for a printer in her head appeared to be linked to what Rose saw as the administrative demands placed upon her rather than what she believed was professionally necessary. For example, she explained the competing discourses when she said

we have appraisals and...one thing they look at is our assessing, our collection of data. Um, like for a running record, we have a minimum that we have to do for the year. We have to take story and handwriting samples every term and it's expected that we have what we call a roll book...where we are collecting data on an ongoing basis on units done and on reading, maths, story writing. So it is expected we cover all areas frequently.

But she also stated

I don't feel I need to [record] all the time. I think if someone came into your room to talk to you about a specific child...you can be specific about where they are at all levels within reason. It really doesn't matter whether it's all written down. If a parent comes in, like you don't run for your data. I'm a parent as well, and, if I go to a teacher and she can only talk about my child

from the data I feel really disappointed. I think that she does not know my child as a person. You should be able to sit and talk about a child and go to the data to be specific but there is far more to that child than data. I think that's the back up, it's not the child.

I like to teach but I don't like to write things down. [In] my ideal school there would be less writing for teachers. We spend a lot of time [writing], planning a unit...planning how you are going to assess it all the time, and sometimes it becomes so overbearing. How much assessment, how important it is.... I realise we do have to collect data especially for report writing at the end of the year. But in the ideal school there would not be the pressure to have to collect all the data.

### Three Instruments of Disciplinary Power

It seemed that while Rose found unit planning, assessment planning and checking each child against the achievement objectives "overbearing", she did this because she was expected to. And she explained that when another teacher (usually more senior than her) appraised her this was an aspect that would be checked. Rose seemed to be explaining how she was disciplined into using the practices that she used. In *Discipline and Punish*, Foucault (1977) explained three "instruments" of disciplinary power: hierarchical observation, normalising judgement, and the examination. I explored this sense of "we have to" further, and Rose told me that

it's a passing-down thing. The Ministry has expectations of the board of trustees and the head of the school and then pressure goes onto us. It just filters down. I think I feel that there is a lot more of that now than before. Yeh, I might be wrong but that's the impression I've got coming back [into teaching after a few years raising children] that everyone's far more accountable right through, all the way down the line.... There were always inspectors, weren't there? I don't know why I feel that it's intensified but it has. I think it's not just in education, it's in all areas now, everyone's more accountable.

In this response Rose was describing the nature of hierarchical observation of teachers within the school. She explained that each level within the school hierarchy would check that the next was performing the functions allotted to it to a

certain standard or norm. She referred to these norms as “expectations”. She explained that the expectations were set out in the school’s policy and communicated in school “lore”.

We have a minimum [of running records] that we have to do for the year. We have to take story and handwriting samples every term and its expected that we have a roll book in the junior school, where we collect data on an ongoing basis on units done and on reading, maths, story writing.... It’s expected that we cover all areas frequently.

Surveillance of the teachers was ongoing in the school through regular appraisal, performance management schemes and regular mechanisms of visits to classes by senior teachers, staff meetings to share ongoing data gathering and checking reports to parents. Through systems set up to monitor the progress and achievement of students, therefore, Rose showed how she was, in fact, herself disciplined. As explained by Foucault,

Discipline “makes” individuals; it is the specific technique of power that regards individuals both as objects and as instruments of its exercise. It is not a triumphant power, which because of its own excess can pride itself on its omnipotence; it is a modest suspicious power, which functions as a calculated, but permanent economy. (Foucault, 1977, p. 170)

Through the combination of the observing hierarchy and its normalising judgement Rose felt herself to be under constant examination and, being so, put pressure on herself to conform.

You want to make a good impression. You want to be seen to be doing the job well. I know that in this school there are quite high expectations. It’s worth it. My children go here. But on the other side of the coin we’re doing the work as well, so...I may put the pressure on myself. I don’t know. That’s what I think. We don’t know how much other people are collecting. I could find out that I’m doing a fine job and that I’m doing plenty but you’re too scared to ask anyone.

From Rose’s responses in the interviews it appeared that a relation of surveillance which was predefined and regulated was, indeed, inscribed at the very

heart of her teaching, “not as an additional or adjacent part, but as a mechanism that is inherent to it and which increases its efficiency” (Foucault, 1977, p. 176).

Clandinin and Connelly (1995), in introducing their work on narrative research with teachers, describe how a “rhetoric of conclusions is packaged and transmitted via the conduit to the teachers’ professional knowledge landscape” (p. 9). While not wishing to imply that the only kind of theoretical knowledge in the conduit is a rhetoric of conclusions, Clandinin and Connelly state that a big part of what is fed down such a conduit is a rhetoric of conclusions, in Rose’s case a discourse about recording data about learning.

From the interview data it appeared that Rose was very focused on summative assessment that could be used to check children’s achievement against the achievement objectives. In order to investigate how Rose’s practice matched her description, I observed her in action in her classroom for a week in March 1998. As explained in the introduction to this chapter, Rose agreed to wear a microphone and tape recorder to capture her interactions throughout the day, and to have me shadow her for the week. It would take a book in itself to recount just this one week’s teaching, so representative examples have been used as a basis for describing Rose’s approach to assessment in her classroom.

After a settling-in period, in which several of Rose’s students asked about the microphone, wire and bum bag in which the tape recorder was housed, Rose and the children in her class seemed to forget the equipment was there and appeared to accept me. There were times when Rose made aside comments to me during the action of teaching. Most of these were also recorded on the tape and provided useful information about her knowledge of children and reasons she took some actions over others. But most of the tape recordings were just daily classroom life.

From the moment she began teaching on the Monday, this classroom was “action packed”. Rose’s class was full of 30 lively 6- and 7-year-olds. The day began as the bell rang and the children settled onto the “mat” (a carpet square in a central position in the classroom). Rose called the roll and chatted with each child as she did so before moving the class outside for an active fitness session. Class

news, followed by a story and then frog-making activities comprised the class programme until morning tea. After a 20 minute break the class resumed with a reading lesson and then a swimming lesson until lunchtime. In the afternoon the class had a maths session focusing on numeracy, followed by a science lesson observing and recording the growth of their class tadpoles. The day culminated in the preparation and presentation of a card to farewell a class member who was leaving the school that day.

The following days proceeded much as the first, except that on Friday morning the class went with the rest of the school to the city swimming pool for their annual swimming sports. I observed Rose interacting with the children, often feeding back information about their behaviour and reminding them about organisational details, such as which group they were in and what activities they should be undertaking. Rose often had information about each child in mind and used this in framing her feedback to children. For example, she complimented one boy on his news reading and helped him expand on the information about conserving lizards in New Zealand. Although from an outsider's perspective this child had struggled through the short newspaper article, Rose explained, in an aside to me, that he was having special assistance with his reading in a reading recovery programme, yet, at age six, he was able to read articles of interest straight from the newspaper. Despite the information that Rose held about each student in her head, she only recorded information about learning on two occasions during the first day's observation.

### Summative Assessments Recorded

The first recording occasion was during the news session mentioned above, when three children presented prepared statements about current events items to the rest of the class. Rose sat in a small chair to the side of the class and prompted and supported each speaker. As they were speaking she recorded in her roll book. Figure 8.1 shows the page on which she was recording information about the children's ability to give the news. She was recording, specifically, their ability to speak clearly and use appropriate volume in their presentation. The code at the top

of the page shows that she was judging the children as below, at or above expectation and, at times, she recorded a brief comment to amplify these coded marks.

The second occasion on which Rose recorded information was during mathematics in the afternoon, when she checked each of the children's maths books to see how they had managed the examples. She placed colourful stickers on some children's work. She didn't write any information in her own records about this work, however.

For the swimming lesson and the reading lesson Rose referred to her roll book to check which groups the children should be working in. For example, see Figure 8.2, which shows the reading page from Rose's roll book. The abbreviated colour names and numbers on this page refer to the level of books at which each child was achieving in February that year. In most cases Rose had added a comment about how well each child was managing. In the far right-hand column are the accuracy and self-correction scores from the running records taken at the appropriate level in February. Rose commented to me that she only recorded which swimming groups the children were in at this time of the year, as she would work specifically on swimming outcomes later in the year.

As Figures 8.1 and 8.2 indicate, Rose was recording summative achievement at certain points in time during the year. As she explained in the interviews, nearly everything she wrote down (and identified as assessment) was summative in nature and used mainly for reporting on progress or to demonstrate to her colleagues and superiors that she was keeping evidence of learning. The headings in her roll book were mostly from the achievement objectives in the curriculum, except in reading, where more specific levels (the colour wheel from the Ready to Read series<sup>24</sup>) were employed.

---

<sup>24</sup> The Ready to Read series is a series of graded reading books produced by a government printing firm in New Zealand for use in teaching children to read. The colour wheel is printed on the back of each book and indicates the difficulty level of the book.

|  | Tell stories, recite or read aloud, informally and for an audience | Converse, ask questions and talk about events and personal experiences in a group | Listen to and interact with others in a group or class discussion | Current Events Unit |                      | TERM 1 SPEAKING<br>+ exceeds expectations<br>✓ meets expectations<br>• did not meet criteria<br>Comment |
|--|--------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------|---------------------|----------------------|---------------------------------------------------------------------------------------------------------|
|  |                                                                    |                                                                                   |                                                                   | 1. Speak clearly    | 2. Speak with volume |                                                                                                         |
|  |                                                                    | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Spoke very clearly & confidently                                                                        |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Spoke confidently - interesting content                                                                 |
|  | ✓                                                                  | ✓/+                                                                               | ✓                                                                 | ✓                   | ✓/-                  | Spoke quietly - very nervous                                                                            |
|  |                                                                    | ✓                                                                                 | ✓                                                                 | ✓                   | ✓/-                  |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Spoke of trip to England                                                                                |
|  |                                                                    | ✓                                                                                 | ✓                                                                 |                     |                      |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Unsure of content                                                                                       |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    |                                                                                                         |
|  |                                                                    |                                                                                   |                                                                   |                     |                      |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Children found item very interesting                                                                    |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    | Quiet voice, but spoke clearly                                                                          |
|  | ✓                                                                  | •                                                                                 | ✓                                                                 | +                   | ✓                    | Spoke very clearly                                                                                      |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | +                   | +                    | Very confident & clear - plenty of volume                                                               |
|  | ✓                                                                  | •                                                                                 | ✓                                                                 | ✓/+                 | ✓/+                  | Spoke..                                                                                                 |
|  |                                                                    | +                                                                                 | +                                                                 | ✓                   | ✓                    |                                                                                                         |
|  | ✓                                                                  | ✓                                                                                 | ✓                                                                 | ✓                   | ✓                    |                                                                                                         |
|  | ✓                                                                  | •                                                                                 | ✓                                                                 |                     |                      |                                                                                                         |
|  | ✓                                                                  | +                                                                                 | +                                                                 |                     |                      |                                                                                                         |

Figure 8.1 Extract from Rose's roll book: Term 1, Speaking

| Reading Term 1<br>(February) | Level of<br>Colour wheel<br>tested | Groups determined<br>from February Running Records<br>and comment. | Accuracy and<br>Self correction<br>rates. |
|------------------------------|------------------------------------|--------------------------------------------------------------------|-------------------------------------------|
|                              | blue 3 / Gr 1<br>ESL / Emergent    | S.C. constantly reruns, monitor for sense <sup>(B3)</sup>          | 97% S.C. 1:2                              |
|                              | Turq. 1.                           | R. for meaning, monitored most errors <sup>(Tq1)</sup>             | 96% S.C. 1:2                              |
|                              | Turq. 1<br>Blue 2                  | fluent, gd. understanding <sup>(O2)</sup>                          | 97% S.C. —                                |
|                              | Turq. 1.                           | fluent, silent dr. - gd. understanding <sup>(O2)</sup>             | 98% S.C. —                                |
|                              | Turq. 2                            | R. for meaning, monitored all rdg <sup>(Tq1)</sup>                 | 99% S.C. 1:1                              |
|                              | 8-8½<br>Gr. 1.                     | 8-9 yrs. fluent, v. gd. comprehension                              | 99.5% S.C. 1:2                            |
|                              | Orange 2 / Turq. 1<br>Blue 3       | integrating all strategies <sup>(Tq1)</sup><br>not rdg for sense   | 96% S.C. 1:2<br>94% S.C. 1:2.5            |
|                              | Orange 1<br>Orange 2 / Gr. 3       | rdg for meaning at <sup>(G2)</sup>                                 | 98% 1:2                                   |
|                              | Blue 2                             | Needs more fluency - good strat                                    | 94% S.C. 1:8                              |
|                              | Gold 2 ✓<br>Gr 1 / Blue 3          | V. fluent, good understanding <sup>(G2)</sup>                      | 99%                                       |
|                              | Green 3                            | Needs to monitor for sense <sup>(G3)</sup>                         | 94% S.C. 1:9                              |
|                              | Turq. 1                            | V. fluent, v. gd. understanding <sup>(O2)</sup>                    | 99% S.C. 1:5                              |
|                              | Turq. 2                            |                                                                    |                                           |
|                              | Gold 2                             |                                                                    |                                           |
|                              | Turq. 1                            | errors - names, v. good comprehension <sup>(O2)</sup>              | 97%                                       |
|                              | Turq. 1                            | very fluent, gd. understanding <sup>(O2)</sup>                     | 99% S.C. 1:1.5                            |
|                              | Purp 1                             | Rd with expression, disc. understand fully <sup>(P1)</sup>         | 98% S.C. n.a.                             |
|                              | Turq. 1                            | good understanding - monitors rdg <sup>(O2)</sup>                  | 97% S.C. 1:4                              |
|                              | Turq. 2                            | fluent                                                             | 100%                                      |
|                              | Gold 2 ✓                           | fluent, but deeper meaning needed <sup>(G2)</sup>                  | 100%                                      |
|                              | Gold 2 ✓                           | not v. fluent, but v. gd. understanding <sup>(G2)</sup>            | 97%                                       |
|                              | 8-8½                               | 8-9 yrs. v. fluent / expressive, v. gd. comp                       | 99% S.C. —                                |
|                              | 8½ - 9                             | 8-9 yrs. v. gd. understanding                                      | 99.5% S.C. 1:2                            |
|                              | Purp 1                             | Needs to monitor sense / rd for deeper meaning <sup>(P1)</sup>     | 96% S.C. 1:4                              |

Figure 8.2 Reading records from Rose's roll book



The reading records, in contrast to the other records, contained specific information related to reading progress. They also indicated what needed improvement in many cases. As well, Rose used this information to form reading groups so that she pitched reading materials and activities appropriately for the children in her class.

### Assessment for Formative Purposes

As well as using her recorded assessment, Rose used her intimate knowledge of the children as she interacted with them individually in every learning area during the day. She kept a great deal of information about each child's progress in her head and compared it with her expectations for children at this age. On Tuesday morning many of the children were still completing the frog models and paintings they had begun on Monday. Others were writing a set of instructions about how to make the frog they had completed. Rose was assisting some to complete their origami frogs as well as checking up on and helping the others to write their instructions. This frog making and writing instructions were part of a unit Rose had planned to assist the children to develop their procedural writing. She had reminded the children about writing instructions before they began making the frogs on Monday, and on Tuesday had prepared sheets for each group on which the instructions had been started. For example, one began, "To make a frog painting you need..." Rose also talked through each activity with the whole group and wrote words they might need, like "paper", "paint" and "paintbrush" on the whiteboard at the front of the room. She reminded them that "frog" and other frog words were on the frog table. For some time at the beginning of the lesson she continued to help children make the origami frogs as they had proved too difficult for many children the day before. The rest of the children moved to their tables to begin writing the instructions.

As the children wrote, Rose moved around, checking they were on task and giving feedback about the behaviour at the tables. Some children approached her with their efforts and Rose reminded them to think about the steps they had taken in making their frogs. She asked them what they had needed to make their frogs in

order to help them write their instructions. She helped some to spell words they were unsure of, but, with others, she expected more independence.

To make a frog painting you need...? There are instructions and words on the board. I'm expecting you to try it by yourself. Look, here are the words, and frog is on the topic board. Come back to me when you've finished. Keep your lines straight. This table is working very nicely, keep it up. I'll be there to help you in a minute. Chris...are you all right? Come and sit here by me. Try and keep your lines straight, are they straight?

It's not quite straight. What happens if you haven't got it straight?

Well it won't come out quite as good.

Good, did you have to have instructions to make the jumping frog? Yes, good girl Anna. Did you mix any paints to get some colours? Did you do an outline first? You go and have a think of all the steps that you used. And now...[reading] "I mix the paint and then I paint". What did you paint? What did you paint first?

[Anna] A frog. Its head.

OK, you need to say that, remember.

In an aside to me during the lesson Rose commented

This is one of my better, you know, top ones, so you can see the range through to Chris, who was really just copying the words off the board. And that just means so much one to one, it's quite frustrating. I've got quite a few Year 2/3s but I've got about six who are not quite as bad as that, but are definitely causing me concern. I've asked Torie [the senior teacher] to come in and work with that group, because I want her to be aware that at the start of the year...I think quite a few of these kids are coming through and its not recorded correctly, and they're having this much trouble. And so it gets to Year 3 and then I worry that I've got to send them to the senior school.

Rose's specific feedback to Chris and Anna, and her comments to me, indicated that she was adjusting the difficulty of the task to suit the needs of each child as they worked. As reviewed in Chapter Four, Bell and Cowie's (1997) model of formative assessment in science explains two types of formative assessment, one as planned formative assessment that is used by the teacher to find out how children are understanding or performing in a particular aspect of

curriculum, and the other as interactive assessment where the teacher notices and responds to learners less formally. I found this model very useful when examining how Rose, and the other teachers I observed, used assessment as a process “to recognise and respond to student learning in order to enhance that learning, during the learning” (Bell & Cowie, 1997, p. 13). Clearly Rose was using interactive formative assessment as she noticed, recognised and responded to each child according to how he or she was coping with the task. Although she had not planned this activity as a formative assessment activity, Rose was aware of the need to structure the learning according to the needs and strengths of each student. Black and Wiliam (1998b) emphasise this as the most important form of assessment needed to promote learning, but Rose, judging from her responses to questions about assessment in the interview, did not appear to view this behaviour as assessment. In contrast, for Rose, assessment was the judgements she made about whether the children had (or had not) met the achievement objectives.

When the time for the lesson drew to an end Rose called the whole class to attention, collected the finished products and added names if they were missing. On a few children’s work she wrote the children’s dictated sentences and then shared some of the writing with the class. Rose thanked the children for their instructions, reinforcing the sort of writing they had just undertaken. This type of lesson was typical in Rose’s week. She was actively involved, giving feedback as needed and working with the children as the day progressed. Throughout the week the only recording about assessment that I observed were the codes she added as each child gave their current events news. I asked Rose about this pattern and asked her if it was usual. She assured me it was how she normally worked. She found she could record notes about speaking so that she had the evidence but only recorded information about progress on other curriculum areas occasionally, often at the end of a unit, before starting a new topic. For example, she intended to record achievement in her roll book about the children’s ability to write instructions toward the end of her time concentrating on this genre. In nearly all cases she recorded the children’s achievement under the achievement objectives copied directly from level 2 of the curriculum (for example, see Figure 8.1).

These results were consistent with the information Rose had given me during the interviews. She was very aware of the children and their individual strengths and needs, but her recording couldn't keep up with the busyness of the classroom action. Much, if not most, of Rose's information was stored in her head. She operated from day to day from her plans, focusing on the learning activities she had organised and recorded information about learning towards the end of a unit, or, as in the case of the current events, as she judged particular children's performance periodically. In order to meet with school policy and be accountable, she mostly used the achievement objectives (usually straight from the curriculum documents) as checklist headings in her roll book, which was specified for use in the school assessment policy.

### Making Assessment Manageable

Rose also collected the samples specified in her school's policy each term.

Well, specifically we do written samples once a term but then I would be assessing what they wrote more often than that. I don't always do checklists when I mark their work though. I don't always tick, yes they can do this, yes they can do that. But I think each time you're marking you make a few little notes [in your head] and then you're building up a bigger picture so that when you do assess more formally, say twice a term, that you do know that child. I can't see the point of ticking off boxes endlessly though, and when I am assessing at the end of a unit I would have their work there with me as well. I'd have written little things through it, and so just by flicking through I can see where that child was at.

[Mary] So you would carry a lot of information in your head?

Definitely. A lot of people jot it down in places but I probably don't do a lot of jottings, so most of it's probably in my head. But then I often think, like I saw somebody doing something today and I thought I should probably write that down knowing that that's the type of thing that people go and jot down. But then I've got to get up and go over and write it down and you lose your thread of what you're doing. So unless you have ...your planner beside you or.... They said to do field notes in your roll book but I think it has become more formal than that. I know that's what they wanted us to do but the way we've set it

up hasn't really allowed for that. That's something we can look at.

[Mary] I thought you were going to set it up how you wanted to set it up?

Well, we did it in teams. We sort of agreed in teams how we were going to do it and I can already see lots of downfalls. And I can see that what I had in mind, how I should evaluate wasn't necessarily even how somebody in my team should evaluate. I can see that it has to be more individual and yet they wanted uniformity in it as well. I think a lot of the curriculum statements have to go into the roll book. I think that's what we're teaching to...so I've cut them out and I've just stuck them in, like even if they're just mind joggers. At level 2. So I'll assess on that. Can this group do blends, is this group able to use a contents page?

As argued in Chapter Seven, Rose was using "grids of specification" (Foucault, 1981) as a way of knowing her students in order to comply with the school requirement to use her roll book to "monitor student progress against the national achievement objectives" (Ministry of Education, 1994, p. 11). Although she stated that she felt each teacher should have his or her own way of meeting that requirement, she saw it as expedient to photocopy and cut up the objectives straight from the curriculum documents in the same way as everyone else in her team had. She also often used "we" in her explanations, indicating the productive nature of team decision making in her practice. Thus, both manageability and peer surveillance were competing with Rose's constructivist ideals here. This division of the curriculum for checking off seemed to be the pattern for most of the curriculum except in reading, and in maths, where one staff member had devised school-wide objectives within the Beginning School Mathematics<sup>25</sup> programme. Rose said

I'm probably checking off a lot more in maths. We've just come up with this new thing, like this is our overview for the year, the topics and our objectives. Anne [another staff member] did them

---

<sup>25</sup> Beginning School Mathematics was a programme produced by the Department of Education for use in junior school classrooms.

up. So we've been copying those in and then as I go I am checking them off more frequently.

Figure 8.3 shows the grid that Rose had in her roll book for the maths unit she was teaching on number/algebra at the time. She used the same key for recording whether the children had exceeded expectation, had met the criteria or had not achieved each criteria as she was using in speaking (see Figure 8.1) and handwriting. However, because the mathematics curriculum had been divided into topics and then each topic into a large number of criteria, this meant Rose felt she was constantly checking off in mathematics. She explained to me that this tended to drive her mathematics planning, as she set out to cover and assess a certain number of the criteria each week. Due to this, the expectation that she check each child against each criterion drove her unit planning as well as her long-term plan to get through the mathematics curriculum. Through the definition of the curriculum as a series of criteria (or standards), and through constant surveillance of individuals using the checklists, Rose and her fellow teachers had produced a mechanism of “normalizing judgement” (Foucault, 1977, p. 184). In detail Foucault explains that the regime of disciplinary power

brings five quite distinct operations into play: it refers individual actions to a whole that is at once a field of comparison, a space of differentiation and the principle of a rule to be followed. It differentiates individuals from one another, in terms of the following overall rule: that the rule be made to function as a minimal threshold, as an average to be respected or as an optimum towards which one must move. It measures in quantitative terms and hierarchizes in terms of value the abilities, the level, the “nature” of individuals. It introduces through this “value-giving” measure, the constraint of a conformity that must be achieved. Lastly, it traces the limit that will define difference in relation to all other differences, the external frontier of the abnormal.... The perpetual penalty that traverses all points and supervises every instant in the disciplinary institutions compares, differentiates, hierarchizes, homogenizes, excludes. In short, it *normalizes*. (Foucault, 1977, pp. 182-3)

| NUMBER/ALGEBRA |                       |                                 |   |
|----------------|-----------------------|---------------------------------|---|
| Objectives     |                       |                                 |   |
|                | Can rote count to 100 | Can record 2-digit numbers. w/c |   |
| Jane           | ✓                     | ✓                               | • |
| Chris          | ✓                     | ✓                               | • |
| Ben            | ✓                     | ✓                               | • |
| Luke           | ✓                     | +                               | ✓ |
| Alice          | ✓                     | ✓                               | ✓ |
| Clare          | ✓                     | ✓                               | • |
| Tracy          | ✓                     | ✓                               | • |
| Melinda        | ✓                     | ✓                               | • |
| Jim            | ✓                     | ?                               | • |
| Sian           | ✓                     | ✓                               | ✓ |
| Alec           | ✓                     | ✓                               | ✓ |
| Vicky          | ✓                     | ✓                               | ✓ |
| Kahu           | ✓                     | ✓                               | ✓ |
| Mai            | ✓                     | +                               | • |
| Tom            | ✓                     | ✓                               | ✓ |
| Mark           | ✓                     | +                               | • |
| Daniel         | ✓                     | +                               | ✓ |
| Toby           | ✓                     | ✓                               | ✓ |

Figure 8.3 Rose's Checklist for her Algebra Unit

In addition to these checklists, Rose kept samples of the children's work but told me that there were too many to send on to the next teacher and that it would be nice to hand them back to the children at the end of the year. The samples were mainly used to say whether or not children had achieved certain objectives and this information was recorded on the checklists. They were collected every three or four weeks.

Depending on what your unit is. But I would probably have a sample of work every three or four weeks at least from one of those main areas.

Through the collection of work samples and the constant completion of the checklists, Rose had sought to meet the syndicate team's expectations, comply with school policy and manage such a large amount of data. But knowing the children through the criteria in this way produced a largely summative bank of information, used only occasionally for grouping children or for reporting in terms of various areas of learning.

### Competing Discourses of Assessment

Trying to understand how Rose made her decisions about what the next teaching steps for her children were, I asked how it was that she knew what a student or group might need. In response Rose told me

I think in our head we have a sort of map and a sort of time line. When I first came back to teaching I hadn't taught in the junior school so I didn't have a long-term vision, so to speak. And I found it really difficult. Every Sunday night was sort of a nightmare, what will happen this week? But now I can sort of see how they progress and where they might be starting from and where I'd like to see them go. So its just another "in your head situation". I don't know where it comes from, but it's just amazing really. Because it is, it's all in there. I can see a picture now and I presume that's what a lot of teachers do because how can you plan if you don't have a forward vision of where you want them to go?

[Mary] Do you go back to the curriculum documents for information?



No, no not all. I don't know whether we're supposed to. Sometimes I do, I mean, some of the terminology. I think, oh what do they mean, what do they want us to do there? Especially with new things like "poetic" where we all think of the old poems and things like that. So I have gone to the curriculums but not regularly. I've tried out a couple of the examples they give and so that gives me confidence,...but I don't get them out very often. I know that ultimately that's what we have to assess to.

The use of "they" and "them" in opposition to "we" and "us" implied a tension between an imposed set of criteria and those she kept in her head. Rose was struggling to "know" what official policies meant in terms of the knowledge she held in her head about learning in particular areas. "They" and "them" referred to the Ministry of Education, the curriculum documents and other official agencies, which sent new terminology "down the conduit" (Clandinin & Connelly, 1995). Clandinin and Connelly used the term "conduit" to encapsulate the way in which an administrator in their research had explained how official knowledge travelled to teachers "via a metaphor of a funnel, which, he claimed, board officials used to pour different policies into different schools" (Clandinin & Connelly, 1995, p. 3). In the self-managing school context, Rose was disciplined by herself, her peers and the knowledge that they would self-review and be reviewed by the ERO to check whether the official knowledge was being taught. The weight and power of these official discourses of assessment often seemed to overpower Rose's personal map and time line about learning – her formative assessment practices – even though she knew these were critical to successful learning. Explaining how disciplinary power works, Foucault (1977) described the school as an apparatus of uninterrupted examination in which the economy of visibility was transformed into the exercise of power.

In discipline, it is the subjects who have to be seen. Their visibility assures the hold of power that is exercised over them. It is the fact of constantly being seen, of being able always to be seen, that maintains the disciplined individual in his subjection. And the examination is the technique by which power, instead of emitting signs of its potency, instead of imposing its mark on its subjects, holds them in a mechanism of objectification. In this

space of domination, disciplinary power manifests its potency, essentially, by arranging objects. The examination is, as it were, the ceremony of objectification. (Foucault, 1977, p. 187)

Rose used the official terminology of the achievement objectives and learning outcomes derived from the strands within the curriculum to document the children's learning because that's what she knew that she would be checked up on. Operating as a self-managing school, the teachers at Arawhata had developed their own codes for measuring the students against the official criteria, and, in doing so, disciplined themselves.

At this school we do... nothing if they haven't achieved it, tick if they're doing it and a cross if they do it very well and I don't put crosses easily. Our focus this year is speaking in [the] language [curriculum] across the whole school, we've set that up for the whole year. And there are written language objectives. Torie did that for the whole team. These are nuts and bolts type things, like... full stops and wide vocabulary sentences. And we've put all these lovely little objectives, but are we going to use these every term?

The tension between what the team had decided to do and the way Rose actually went about her assessment seemed to be a factor in Rose's frustration with assessment, especially the recording side of it – needing a printer in her head. While she worked formatively, assessing against the “picture” she held in her head, Rose knew that she was expected to check progress against “these lovely little objectives”, which didn't often seem to tie in with her teaching. She got around this by taking samples of work from the children periodically to help her fill in her checklist. But she was still left wondering if she was doing enough, and finding it very time consuming.

I think it will always be time consuming. I don't mind. I see it as part of the job, but it is quite time consuming and sometimes it accumulates on me and that's where I worry. You know I did something and I assessed it, but you should really be writing it down within a day or two but it doesn't always happen that way. It's put to the bottom of the basket and then you might have a writing sample and you should be making all your comments that day, before school the next day, but it just doesn't happen. But we can't be perfect!

This suggests that there was a clash between what she believed was the way she should be assessing and her ability to do so. It appeared that the summative, checklisting discourse agreed to by all the teachers in her syndicate was what Rose believed she should be doing. But she had found that operating that way was not often what happened. This left her with a sense of not being perfect, of not living up to her own expectations of herself, and, consequently, feeling anxious and out of control. In order to cope, Rose used summative samples and periodic tasks and running records to keep track of children's achievement in terms of the achievement objectives. Hence, she appeared to be juggling a dual system of assessment, one in which the interactive, formative aspect rarely came in contact with the summative one in practice. This could explain why Rose found her assessment time consuming yet still worried that she might not be doing enough.

Rose's daily practice was formative, mainly interactive (Bell & Cowie, 1997), where she noticed, recognised and responded to learning as it occurred within the programme. She did not appear to plan more formal opportunities to gather information that she could act on formatively. There were no clear links between her summative checklisting strategies and the interactive formative assessment. A consequence of this seemed to be that Rose was struggling to keep her assessment manageable as she held one map of progress in her head and recorded achievement against the curriculum from time to time against another set of objectives. The managerial standards-based approach to constructing the roll book set up with a grid of achievement objectives almost all copied directly from the curriculum documents decided upon by the syndicate assumed importance for Rose. She used this approach to plan units and to check off achievement. But in her daily practice Rose mainly worked by matching her mental image and time line of what progress in each curriculum might be with where she felt each child was performing at the time. This suggested that Rose believed that children learn at their own rate and need ongoing personal assistance and feedback to make progress. Rose worked in what could be called an activity-based way, setting up

learning centres and interacting with individual children to assist them to learn to write instructions for frog making.

Within this self-managing school, Rose's assessment practices had developed in two directions. Rose appeared to be responding to both the accountability and formative discourses of assessment identified in Chapter Six. In terms of accountability discourses of assessment, she clearly felt a need to keep to the team decision of her syndicate to track achievement using her roll book as a checklist containing the curriculum achievement objectives. She was meeting the syndicate policy to keep these records as evidence of her accountability to government and school requirements by gathering evidence periodically, usually at the end of units of work. Although she recognised that there were weaknesses in this approach, Rose seemed committed to continuing this approach as long as her syndicate and school policy deemed it appropriate. Thus the NAGs, the school's policy, and her syndicate's approach to meeting them were very powerful influences on Rose's teaching practice.

But Rose also believed it was very important to be able to report and talk to parents about the achievement and progress of students in her class. In fact she stated that she felt it was very important to be able to talk about children's learning in detail, without notes, if approached by a parent. In order to do this, and to support learning, Rose explained that she had developed a mental picture of what progress in certain curriculum areas looked like over time and used this (in contrast with the achievement objectives) in her day-to-day work with her class. The weight and power of official requirements, review and audit regulations, and school management practices had disciplined Rose into keeping written records about whether certain objectives had been attained by the children. These school management practices included hierarchical observation by colleagues, the expectations of the assessment policy and periodic surveillance of her assessment records and reports on children in terms of the achievement objectives. In addition, the prevailing managerial discourses of self-management (Fitzsimons, 1996) influenced Rose to monitor and manage herself to keep up with the amount

of monitoring and recording of achievement she felt she should be doing in order to do her job well.

But in addition to these accountability reasons for assessment, Rose had developed a mental picture and time line that she used constantly to make judgements about children's learning and to make teaching decisions. Although associated with the achievement objectives in the curriculum, Rose's mental progress map was far more specific and allowed her to calibrate learning for individuals. In order to accommodate her use of her personal progress maps and to check achievement against the achievement objectives in the curriculum, Rose resisted recording assessment information continuously during teaching, leaving it until the end of units or periods of time. In this way (although she talked about being continually anxious about whether she was doing enough assessment) Rose made her teaching manageable.

## **Luke**

When I first met Luke he was in his third year of teaching. At that time he was teaching a class of Year 4/5 children (8/10-year-olds) at Waimaka School. His responses to the questions in the interviews led me to believe that he kept most of his assessment information in his head, testing the children from time to time in order to gather evidence of their learning for recording and reporting purposes. I selected Luke as a teacher to observe because he represented the group of teachers who explained their approach to assessment in a way that led me to describe them as "head noters".

Luke's responses to the interview questions indicated that there were few assessment expectations at Waimaka School. There was no written policy on assessment when I first visited him, and Luke explained that they were between two systems of assessment due to a change of principal, although they were still expected to complete the "pink cards"<sup>26</sup> half yearly. Roll books had been

---

<sup>26</sup> "Pink cards" were the Keeping School Records cards (see Chapter Two), which had been produced and distributed by the Department of Education prior to 1989 and were still being used by

distributed for use by each teacher, and each child had been asked to purchase a scrapbook or folder to be used as a sample book. Examples of a child's work in various curriculum areas were collected in his or her "sample book", which was sent home for parents to see progress throughout the year. Luke expressed the expectations thus

We've got the roll books. There's an observation kind of record book too and we're supplied with both of those. It wasn't really said that we should use one or the other, there's a choice. I have my roll book and that's OK. I'm changing over slowly. She [the principal] wanted comments, so I just had to expect that there should be some more comments rather than just ticks.

### "I'm a bit old fashioned": Using Tests for Assessment

Although in his twenties and only a third-year teacher, about his own practice Luke told me

I'm a bit old fashioned. I like an oral or written test, especially in maths. I mean, you can't beat them really, we have spelling tests every week. We do maths pre-tests and post-tests.... I'm not into closed-book tests (in science, social studies), I mean I use worksheets and stuff.

Luke used the results of the tests to track children's progress, particularly in spelling or times tables in maths (see Figure 8.4). He used running records to assess reading achievement once a term (see Figure 8.5), the Progressive Achievement Tests (see Figure 8.6), and recorded this information in his roll book. But day to day, Luke explained, he kept most of his assessment information in his head and used things he noticed about the children's work to alert him to misunderstandings or learning needs.

I always check.... They might not understand the question. Like for homework we did an ordering thing where they had ten numbers to put in order. Some did it from smallest to biggest. Some did biggest to smallest. They were all hundreds numbers and some ordered the digits within the numbers. So that's a

---

some schools as cumulative records of the individual progress and achievement of primary students in the 1990s.

classic example. I had an image in my head of the correct answer and if I had used that to record achievement then that's bad assessment because there were really three possible "right" answers.

By the time I approached Luke about observing him in the classroom he had changed schools and was teaching in the senior class of a two-teacher school in rural Waikato. The 20 children in his class ranged from Year 3 to Year 6 (7/11-year-olds). Because of the wide age, and therefore ability, range of children in this class, Luke tended to work with the children in groups. The daily timetable in Luke's room followed a regular pattern while I was there, beginning with fitness and assembly each morning. Maths took place before morning interval, which was followed by buddy reading with each child of the junior class paired up with one from Luke's class. Reading and English activities including writing occupied the rest of each morning. In the afternoons the class was working on a water safety unit and had swimming lessons in the school pool.

#### Planned Formative Assessment: Using Test Information in Teaching

During my first day's observation, Luke's class took part in assembly with the other class before beginning the day's work with a maths lesson. Luke introduced the lesson by giving out some certificates to children who had done well in the previous week's maths test. He set the older children to work on some maths games and gathered the Years 3 and 4 children together with their maths books. He had the test papers from the previous week's pre-test on number. He sent away the two children who had everything in the test correct to work on other activities.





| Group | 7-8        | 8-9        | 9 $\frac{1}{2}$ -9 $\frac{1}{2}$ | 9-10       | 9 $\frac{1}{2}$ -10 $\frac{1}{2}$ | 10-12      | 11-13 | 12-14 | 13-16 |
|-------|------------|------------|----------------------------------|------------|-----------------------------------|------------|-------|-------|-------|
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| V     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| T/V   | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| R     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| V     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| C     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| R     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| R     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| K     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| C     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |
| C     | [REDACTED] | [REDACTED] | [REDACTED]                       | [REDACTED] | [REDACTED]                        | [REDACTED] |       |       |       |

orange

| Names (removed) | A G E |    |    |               |    |          | C L A S S |    |    |    |    |          | Rdg Age   |
|-----------------|-------|----|----|---------------|----|----------|-----------|----|----|----|----|----------|-----------|
|                 | RC    | RV | LC | SS (Ref. Mat) | M  | Ind. Age | RC        | RV | LC | SS | M  | Ind. Age |           |
|                 | 63    | 55 | 91 | -             | 68 | 69       | 63        | 50 | 88 | -  | 67 | 67       | 4-10 9    |
|                 | 72    | 22 | 63 | -             | 71 | 57       | 75        | 43 | 68 | -  | 79 | 66       | 8½-9½ 10  |
|                 | 18    | 2  | 57 | -             | 27 | 26       | 12        | 3  | 53 | -  | 27 | 24       | 7-8       |
|                 |       |    |    |               |    |          |           |    |    |    |    |          |           |
|                 | 36    | 49 | 58 | 64            | 62 | 54       | 33        | 47 | 58 | 70 | 61 | 54       | 12-14 8½  |
|                 | 42    | 42 | 76 | 30            | 16 | 41       | 23        | 35 | 64 | 32 | 13 | 33       | 8½-9½ 8½  |
|                 |       |    |    |               |    |          |           |    |    |    |    |          |           |
|                 | 40    | 48 | 56 | 51            | 24 | 44       | 42        | 53 | 58 | 39 | 28 | 44       | 10-12 9-  |
|                 | 75    | 74 | 81 | 81            | 84 |          | 74        | 74 | 80 | 82 | 82 | 78       | 13+ 10    |
|                 | 36    | 39 | 10 | 12            | 22 |          | 33        | 35 | 8  | 15 | 19 | 22       | 11-13 1½  |
|                 | 77    | 74 | 98 | 84            | 55 |          | 77        | 74 | 97 | 86 | 57 | 78       | 11-13 11  |
|                 |       |    |    |               |    |          |           |    |    |    |    |          |           |
|                 | 67    | 38 | 43 | 51            | 37 |          | 63        | 33 | 41 | 45 | 35 | 43       | 10-12 11- |
|                 | 36    | 17 | 4  | 8             | 13 |          | 30        | 13 | 3  | 7  | 12 | 13       | 10-12 9-  |
|                 | 54    | 69 | 76 | 51            | 31 |          | 49        | 65 | 74 | 49 | 29 | 53       | 11-13 10- |
|                 |       |    |    |               |    |          |           |    |    |    |    |          |           |
|                 | 26    | 34 | 5  | 18            | 10 |          | 24        | 30 | 4  | 19 | 12 | 18       | 10-12 8½  |
|                 | 79    | 80 | 74 | 84            | 68 |          | 66        | 65 | 59 | 70 | 53 | 66       | 13+ 11-   |
|                 | 77    | 85 | 81 | 91            | 57 |          | 83        | 85 | 79 | 90 | 65 | 80       | 13+ 12    |
|                 |       |    |    |               |    |          |           |    |    |    |    |          |           |
| Part 2 Ave      | 51    | 26 | 70 | -             | 55 | 51       | 50        | 32 | 70 | -  | 58 | 62       |           |
| Part 3 Ave      | 61    |    | 63 | 54            | 44 |          | 47        | 53 | 61 | 54 | 43 | 62       |           |
| Part 4 Ave      |       |    | 47 | 51            | 36 |          | 53        |    | 43 | 47 | 34 |          |           |
| Class Ave       |       |    | 58 | 52            | 43 |          | 50        |    | 56 | 40 | 43 |          |           |

Figure 8.6 P.A.T. Results from Luke's Roll Book

Luke then proceeded to go over aspects of the test material that the rest of the group had struggled with, acting on the basis of the assessment information to alter his practice. He focused first on the place, face and total value of certain numerals. Luke wrote the words “face value” on the board and asked the children the face value of several digits in a number he wrote on the board. When all the children in the group had demonstrated they knew what was meant by face value, he wrote “place value” on the board and asked what that meant. Alan answered

It means like if it’s in the tens or ones.

[Luke] Too smart for me aren’t you? So what’s the place value of this number here? What place is it in, Alan? Which one do you choose? Ones or tens?

[Alan] Tens.

[Luke] Are you sure? Why do you say tens?

[Alan] It’s tens. Its higher.

[Luke] It’s higher. So what’s this one here worth then Bruce?

[Bruce] Umm, in the ones?

[Luke] This one here?

[Alan] Tens.

[Luke] So you think it’s the ones and Alan thinks it’s the tens. Which one do you think it is, guys? We’ve got double answers here.

[Bruce] Seventeen.

[Luke] Seventeen. So is this number saying ten and seven is it? All right, what’s that one then? Is it tens or ones?

[Many voices] Tens. Ones....

[Luke] It’s what?

[Dennis] Tens.

[Luke] If that's tens, then...

[Dennis] No, that'll be seventy.

[Luke] Oh, he's onto it. What's that one, is that ones or tens?

[Dennis] I know.

[Luke] I know you know Dennis. I'm trying to see if these guys know. That's the tens, so is that the tens as well then?

[Alan] No.

[Luke] No. Then let's just see if we know what we're talking about. If we're talking about place value it means if its in the ones or the...?

[Bruce] Tens.

[Luke] Tens column. And then we did those things where we did ten plus seven, and twenty plus seven, do you remember that? Are you sure? So what's this one going to be (pointing to 73)?

[Bruce] Seventy plus three.

[Luke] Good. So that's the tens column isn't it? Whose is this one [looking at another test paper]? Kelly, it's yours. I was really impressed with you. All right you can go now. OK Alan I was a bit confused here because you wrote ones, ones, ones. So what's that one going to be?

[Alan] Oh, tens.

[Luke] Good, so you did understand it. Go away, Alan. You can go to the maths games. I just want to go quickly over fractions with the rest of you.

Luke continued to work with the children in small groups, going over the aspects of the test they had appeared to have trouble with, sending some away and calling others to join the group as the lesson proceeded. Later Luke explained to me that he had recorded the results of each part of the test against the achievement objectives on his checklist but that he had had to ask the children he suspected of having difficulty about several examples as he felt the test did not reflect their

maths knowledge and understanding (see Figure 8.7). When he had satisfied himself whether or not they did know about face and place value, he checked these aspects off against their names in the checklist as well.

Luke had planned his assessment as a test, and although he had recorded achievement summatively against the appropriate achievement objectives that the test measured, he had also used the test as a planned formative assessment device (Bell & Cowie, 1997). He used the children's test papers to explore their understanding further, asking questions to elicit additional information, scaffolding learning about the place-value nature of the number system with the children as he went (Dixon-Krauss, 1996). He amended some of his records in light of what the children did and said in these situations. In this way, Luke was simultaneously carrying out assessment for formative and summative purposes.

Rather than using formative and summative purposes to describe classroom assessment, Vygotsky identified formative assessment as dynamic because it measured "the student's assisted performance during collaboration to assess...what the student is in the process of learning" (Dixon-Krauss, 1996, p. 125). In contrast, "static assessment refers to measuring the student's actual development or what the student has already learned" (Dixon-Krauss, 1996, p. 125). In a vital sense dynamic assessment can relate test results to observations, blurring the distinction between formative and summative assessment, but keeping the formative function uppermost by personalising and using the results.

### Dot, Tick, Two Ticks: Summative Assessment Codes

In such a small school, where Luke was both teacher and acting principal, there were no specific policies about how records should be kept. In his record book Luke used a three- (or four-) point code to record information about learning. As Figure 8.7 shows, his record of the children's learning in mathematics was recorded against their names with a dot (covered but not understood), a tick (covered and understood) or two ticks (mastered). In some cases Luke added a



third tick, which he explained meant they needed extension. He had organised the achievement objectives from the curriculum in levels 2 and 3 and recorded whether each child had achieved each objective through the use of the test and the follow-up discussion. He kept a record, too, of the results of basic addition and multiplication/division facts, also assessed through regular tests throughout the year (as in Figure 8.4).

For the second half of the maths session Luke called all the children to the mat and introduced them to the idea of making patterns with Torro blocks (a construction-type material). He made a few patterns with the blocks to demonstrate the sort of patterns he was on about, then split the class into pairs to build patterns and share them with each other. In an aside Luke commented to me that he used this type of construction material for maths as he had noticed the girls had little experience with spatial construction materials and didn't appear to choose to use them. He also commented that he had thought the place and face value work would have taken longer, but when they had all appeared to have understood it he had changed direction and begun to introduce the next algebra topic with the patterning activity.

Later during that day, Luke asked the children in his class to head up a page in their brag books with the topic "Beach Rules". These were the books in which the children recorded samples of work that were taken home to share with parents how they were doing at school. Luke commented to me that this was the first time he had used the books this way. Another form of home contact was achieved in this class through homework books. Luke marked these daily, keeping a record of who had done their homework. As Luke marked the homework books he called individual children up to his desk to give them specific feedback about how to improve their spelling or how to answer a question in full instead of just giving one-word answers. In one instance he gave positive feedback to a girl, reinforcing the fact that she had researched her answer. He placed stickers on the homework books as he marked them and wrote comments on the books, such as "Good answers – I want to know why you think that."

Although Luke did not record much in his record/roll book, using the code explained above in most cases if he did, he explained that he used his planner to write out units, keep long-term and daily plans and make a note of groups, homework achieved and anything he wanted to recall quickly. He explained that he would do the PATs (Progressive Achievement Tests) next week and these results would be recorded as well. But when asked where he kept the data about students he used for teaching, he said most of it was in his head.

I've only got 20 children so I don't need to have lots of checklists. I can remember 20 children. I've had them all last year and I can compare them with [how they were] last year. To check I can go back and look at their brag books.

### Encouraging Evaluation

Luke recorded the summative test results, but he mainly relied on information held in his head to mediate learning. He gave specific feedback to individual children constantly throughout each day. Towards the end of the second morning that I observed, Luke explained to his class that they were about to attend a water safety presentation given by a visiting drama group. He asked the children to think about the performance and evaluate its worth while it was occurring. He told the children he would like their opinions later on whether they should have the performance group back again, whether it was value for money. In the afternoon, following the performance, he asked for feedback. All the children indicated by raising a hand that they had enjoyed the programme, but one child said that he didn't think it was value for money. This child explained that although it was fun, he felt they could have learnt new things about water safety from it and he didn't feel they had. Luke followed up on this, asking if perhaps they had learnt but were just not aware of it. The children discussed this idea with him and gave examples of things they had learnt while having fun, such as doing up a life jacket firmly and identifying water hazards on the farm. Luke was assisting the children to see that learning is not always work. They spent about 15 minutes evaluating the programme and decided that they would have the group back again, if the occasion arose again.



Luke often built in such opportunities to revise ideas, evaluate learning and programmes and improve self-assessment strategies generally. His use of specific feedback and asking further questions of children infused his teaching. He set up and encouraged peer assessment as well. For example, the children tested each other in pairs on their spelling and tables most days, and through buddy reading more able readers worked with less able ones to assist them to improve through formative feedback. Luke roved round the classroom much of the time, stopping to ask questions and check understanding as he went. In this way formative assessment in Luke's classroom was undertaken by both the teacher and the children (Bell & Cowie, 1997; Sadler, 1989).

Luke regularly monitored learning. Although he did not record a great deal of the information that he gathered about the children, he used it constantly to adjust his teaching, ask further questions and give specific feedback to individuals about their learning. He also used the mental notes he made to re-group the children when necessary and functional. In these ways he was very much a head noter in terms of the way he kept information about learning. He recorded little information in his formal records, using the children's books to jog his memory when reporting to parents or checking up on a particular child's progress. In contrast with Rose's practice of documenting children's achievement in checklists for every unit, and sometimes as she watched them work, Luke's priority was to work to the criteria for learning he held in his head, recording only occasionally. Although he did plan units, he did not stick rigidly to the objectives he had set for them and changed direction when he found the children needed something different. His explanation of his teaching to needs within whatever unit he was teaching spoke of a developmental, child-centred discourse, focused on extending learners rather than covering objectives.

An example of his ability to change direction occurred on the last day I was there. Luke had planned a lesson with the Year 3 children in his class to teach them how to use a calculator. As he worked with them he discovered they could all use it already for adding and subtracting so he changed direction, using the calculator as a tool for the children to show him whether they could work out the

order of three digit-numerals. From this he ascertained that three of the 7-year-olds should be working with a higher group and stored this information mentally for the next week's planning.

Maybe the size of the school and his class meant that Luke was not particularly concerned about documenting assessment information apart from his summarised test results. When describing his practice, he mostly used the pronoun "I" to explain what he did and why. In contrast, Rose often used "we" and "they" to refer to her syndicate team and more senior or official sources. In this small school Luke did not appear to feel the surveillance of others so immediately, and he *was* the principal. He knew in advance when he would be reporting to the board of trustees or the ERO and brought his records up to date periodically for this purpose.

Watching Luke in action allowed some of the gaps in my description of head noters in Chapter Seven to be filled in. Although Luke did not describe himself as using formative assessment, he was constantly noticing, recognising and responding to children's learning (Bell & Cowie, 1997) and adapting his teaching in light of what he observed. As well as this interactive type of formative assessment, Luke set up situations for planned formative assessment (Bell & Cowie, 1997) and used this to obtain information from the whole class as well as groups about their understanding and progress in learning across the curriculum.

### On Becoming a Junior School Teacher

Luke did not appear to feel overloaded by assessment and recording. During the time I observed him in action, he stated that he believed he was keeping enough records to satisfy accountability demands. But a year later he contacted me to ask if I could come out to hear about his new experiences of assessment. A new principal had been appointed and Luke had begun the year teaching the junior class (5/7-year-olds). His first few weeks teaching at this level had been exhausting and he explained that he was experiencing teaching and assessing differently.

There's certainly more planning involved. Being a junior teacher there's just a lot more preparation involved, a lot more planning, a lot more in-depth stuff. With a reading group in a senior room you just go and get them to do a project on tigers or something. You can't do that [here]. You've got to give them activities that keep them – I don't want to use the word "busy" because that's the wrong word – but working actively in short bites, that's the biggest problem I've found.

Luke's description reminded me that all the teachers that I had labelled as "integrated systematic" in their approach to assessment had been junior school teachers. His comments alerted me to the differences between teaching literate and preliterate students.

In those language areas you're monitoring every single individual and you're seeing how because they're growing. I mean in a senior classroom I didn't do that, I didn't monitor individually, if a kid was having trouble you'd talk about it with them and that was it and in that room I was making notes in my head.

Luke's use of monitoring here was interesting. It suggested that he saw monitoring as a written activity. This was consistent with the interpretation of the term by most of the teachers I had interviewed. Monitoring in most cases had meant developing a way of recording whether or not each child had attained a particular learning outcome/achievement objective. For Luke, "making notes in his head" did not appear to be counted as monitoring. In contrast, when he began to teach in the junior school, he believed he was building up a far more detailed data-base on each child, surveilling them more closely against what he believed indicated progress.

I've already got a database worked out with every junior kid. I can tell you what half the kids wrote today in their story and I could never have done that in the senior room, it wasn't so important. I think it's because at the junior level I'm just far more focused in on that, their progress, they are developing. In the senior level you don't measure development in weeks, I mean perhaps you measure it in a year. I mean kids did improve last year, I hope they did. But in the junior room, it's vital. I notice it even more that I'm trying to push them forward and I don't think I had that same push [in the senior room]. That's

why I'm focused into [the fact] that Kerry today wrote two sentences and towards the end he lost it and just wrote words. So tomorrow I'm going to get him to really write more sentences down and read each sentence back to me. And then go and write another sentence, not just go on writing because [that's when] he lost it and he just wrote words and it didn't make sense.

This was an interesting development. Although Luke was still focused on a developmental discourse about learning and teaching, he saw the need to make records about the daily progress of these younger children. He found it more difficult to keep this information in his head and he had had to design a way to record it and use it in his planning. At the time he called me in to tell me about these changes, he had not come to a decision about how he would keep this information and was simply jotting notes as a reminder. This was similar to the way in which the junior teacher I had interviewed and reported on in Chapter Seven as a head noter had coped with the information she gathered about children of this age group.

### Dynamic Assessment in Action

Luke explained how he thought about each student during teaching. He used the example of a girl who seemed to have made a step forward in her learning that morning.

You're focusing in on the kids and what they've done. Today Margy made a breakthrough. She's in Year 3. She wrote such a long story I thought, "your writing is great, now you can go and edit it". No one in Year 3 is into editing. Tomorrow I'm not going to give her so much help, I'm just going to underline them [incorrect spelling] for her and she's got to go away and find them. But that's a huge jump forward and that's why I'm monitoring. Two days ago she was still fumbling around. She was writing better and her stories were getting longer and longer, but today's story was perfect. The sentences were all formed and also there weren't that many words to correct. They were all phonetic words that she could find the corrections for and that's why I chose today. If they'd been words like "charts" I probably wouldn't have said "go and find the words". But today all the words fitted together. She had six or seven minutes left before I stopped everyone else. She could have got a reading book but I

thought “no, get her to do some editing” and she succeeded. I’ll try it again so that’s marvellous.

Luke’s explanation of his own head work shows clearly that he knows each child, how he or she is currently operating in, for example, writing. He also understands the progressions in learning to write. He wanted the children to write more, but also, once they could write in sentences, it was important to encourage them to check the spelling of words for themselves. He explained how he was scaffolding Margy’s learning so that the next steps were achievable. Here again Luke is describing how he holds a mental image of what progress in writing is in his mind as he interacts with individual learners. Simultaneously he compares the learner’s development against his mental progress map, makes decisions about her potential for learning and then scaffolds the next learning step for her to attempt (Vygotsky, 1986). Luke described his view of how learning happens and how he caters for it, saying

I mean, do you agree with Piaget that it’s already stuck in their heads? I mean do they learn from seeing it? [Or] do they learn from hearing it; do they learn from the concrete; do they learn from sitting on a chair [or] from working on the mat; do they learn from rote learning or do they need teacher guidance? It’s huge. Don’t give me questions like that. How do I think children learn? In my head whenever I watch the children work, I don’t set down that they must stay in their seats. If they wish to work sitting in a small group at the table, on the floor, in contortionist positions on the cushions, on the bench in the sunlight that’s fine. Because no child has the same learning needs as any other child and the way they learn is infinitely different.... In the classroom you just have to set up an environment that caters for those different learning needs, perhaps not all at the same moment, but through the year’s programme there will be situations where a child’s particular learning needs will be catered for. Be it in learning centres or copying off the blackboard in spelling, or an activity off a card in reading. So perhaps not every single moment but there will be times when, hopefully, you are getting to that child.

Luke was making it quite clear here that he held a progressive, child-centred view of learning. But he also was not keen to verbalise his tacit theoretical knowledge. He was speaking directly to me when he said “Don’t give me

questions like that!” But then he went on to verbalise that he thought children learnt in infinite ways which he should attempt to cater for. He was realistic in his ability to this, but believed he could influence the learning of all children through providing a varied programme.

### Assessment or Evaluation: Competing Discourses?

For Luke, maybe due to the fact that he was working in a small school and that he had been the acting principal, the need to account to others about the learning of children in his class appeared to be subordinate to his desire to mediate learning in the classroom. He appeared to hold individual children’s needs as a high priority. He did keep records about children’s achievement and he was able to report on the progress that he deemed children were making through the curriculum, but by his own admission he made notes “very occasionally. I know I should do it more...but the battle was getting it onto paper.”

For Luke, most of his assessing was “common sense”. He used the term “evaluation” for formative assessment and added

to me assessing is simply “yes, no, introduced, developing, achieved, or exception” – that’s assessing. I could show you my roll book for that.... That’s assessment – to me it’s cut and dry, do or don’t. And it’s a different form that I do sometimes. That’s assessment

[Mary] OK, as distinct from...?

Evaluation. In my mind evaluation will be comments, using the observation book...in my head I’m doing my evaluations, but I should do them more on paper. That’s what ERO would like. In an evaluation you’re thinking, you’re commenting on it, you’re evaluating the assessments. You’re saying “Rina is at this level, therefore this is where she’s been, this is where she’s going to.” I can see from this evaluation how she’s going, what she’s doing, how she’s working, her learning needs, the theory of learning for that child. If I’m assessing Jack for maths I can say he knows his two times table, developing on the three times. If I was evaluating Jack, I would say Jack is really excited about doing it, he’s really pushing himself and motivated to learn it. He’s doing it at home all the time and to me it’s great news. That’s not

assessment, that's evaluation in my mind and that's what I should put down in the book.

Here Luke is struggling with what he does and what he thinks he should be doing for the ERO. When he is asked whether he makes notes in his observation book (as distinct from his roll book) Luke goes on

Very occasionally, I know that I should do it more that's half the battle, the other half is actually getting it onto paper. I've done some observations, like watching them at learning centres or group work. The assessment is really a little bit of oral language and perhaps visual as well. But really group work is more an evaluation thing. Your evaluating how they work in a group and looking for improvements.

The way that Luke talked about assessment as a battle and compared it with evaluation, which he felt more comfortable with, seemed to indicate that he wrestled with competing discourses of assessment for accountability and evaluation to support learning. By his admission, evaluation usually won hands down in his classroom, with assessment recorded against the achievement objectives relegated to tests and occasional check-ups. Assessment discourses, as "common sense", seemed to have come from official places, such as teachers' college and the ERO. Luke explained

Teachers' college [said it was] simply teach, you assess by looking at the learning objectives and simply assessing is have I achieved my objectives, a measurable objective? So that's what teachers' college taught us...you can't put down my children enjoy working with blocks because how do you assess whether they enjoy it? Even if you ask them, is that valid?

Luke explained that at Auckland Teachers' College (1991-3) there was no such thing as learning needs. It was goals and objectives and "children should be able to". Interestingly, he told me that that was in the first two years but in the third year "it totally changed to...children will be able to." He went on to say that from his memory the evaluation words "walked out the door" and all the "key words were there in the third year, and, of course, that's when the new curriculum documents were coming out". In terms of the ERO's impact he told me

At my first school, we had an ERO visit the year I was there and so assessment was something they picked up on. It was discovered that we weren't uniform on our assessments. So it was something that came up through syndicates and staff meetings. So I learnt that "square" system.

The "square" system was the way Luke had learnt to code whether the children had been introduced to or were developing in, achieving or exceptional in an achievement objective from the curriculum. Although he felt the ERO had an influence, Luke said he didn't see it as being above him. He referred to other teachers as influencing his teaching. For example, he didn't like the way one teacher he had worked with planned to cover a page a day and moved through her plan each day.

When I actually pointed out that this child hadn't learnt yesterday's work, she said "sorry, we have to carry on. There is a failure rate in the classroom, you can't expect to get 100 per cent everyday." I try not to do that.

It was difficult to establish just why Luke thought and taught as he did, why certain assessment discourses predominated over others in his practice. Certainly Luke himself was committed to a child-centred, developmental philosophy. He stated that his life experiences and his interactions with other teachers sustained this approach. In addition, he was teaching in a small school. He had been the acting principal and there was no one other than the board of trustees to report to. Luke appeared to have a good relationship with the board and the community. Many times while I was there members of the community came into the school to help out, do chores, make repairs to the facilities or just catch up with what was happening. By Luke's own admission, he did not see the ERO as above him, but rather as a check to see things were covered and assessed from time to time. While there was documentation, Luke's priority was moving children's learning along a continuum mainly with assistance from the notes he held in his mind. But, at the same time, he tracked children against the achievement objectives. Albeit in a minimalist way, he had constituted for himself



a method of monitoring achievement that he believed met the official requirements and indicated progress.

## **Anne**

Anne is a very experienced teacher. She has taught in the junior part of the school for 15 years and has developed a particular approach to her classroom practice. She has spent weeks at a time, on several occasions, demonstrating her teaching to teachers both in New Zealand and the USA. She has mostly taught in team-teaching situations (open-plan classrooms) at Arawhata School.

The classroom in which I observed Anne was an open-plan space, shared with two other teachers. One of these teachers, Lisa, operated at the opposite end of the room from Anne but shared parts of the programme. A third teacher, Jane, worked part time with some children from Anne's class in the morning. This was to take some of the load off as Anne had over 30 children in her class and was also the senior teacher for the syndicate of Year 1 classes.

Each day in Anne's classroom began as the children arrived at school, in most cases with parents or caregivers. Anne had set up routines which involved the parents/caregivers spending a few minutes settling their children into the room, checking that homework was unpacked and completed, and often interacting briefly with Anne about their child's needs or progress. Either Anne or Lisa began the day with all the children gathered together, leaving the other free for a short time to settle latecomers or spend a little time individually with a particular child. There were also other times during the day when one or the other teacher had short periods of release in which to attend to the diagnostic assessment or other needs of individual students.

Over many years, Anne had developed a set of records that contained the progressions she expected of children of this age across the curriculum and had prepared documents in which this information was continuously recorded. Some were in a book that she generally had with or near her, but others were within the

students' exercise books, on cards, or in the folders the children took home with them every night.

"A circular diagram in my mind: assess, plan, teach"

Anne held the set of progressions she expected most children to move through as they became more competent speakers, readers, writers and mathematicians in her mind as well as in her records. Describing how she went about her classroom assessment in the first interview with me, Anne said

I keep a circular diagram in my mind – assess, plan, teach, assess, plan teach. And as long as it fits into that plan I feel the children are benefitting from it. I've set up a book that contains all my recording and I find that all my objectives I've put are child-centred. They fit thoroughly with what happens with most children at the New Entrant level and the different subject areas. So I've got a focus which becomes part of my weekly planning. If that child hasn't got a particular skill or hasn't achieved something there, in oral language [for example], then that goes into my work plan.

This description of her practice cued me in to observing in the classroom. Anne's explanation and the assessment records she kept, although very detailed, were not motivated by the curriculum documents. She described her objectives as child-centred. She explained to me that they had been derived from her experience over many years of teaching as well as theoretical ideas from Vygotsky, Bruner and Bronfenbrenner. She stated that "scaffolding", based on Cambourne's books, was really important to her. She told me "I provide support as long as necessary until stability is achieved."

Anne's age and length of teaching experience placed her within that group of teachers who were trained at the beginning of the 1970s, when progressive discourses were prevalent and child-centred discourses carried much weight and power. As discussed in Chapter Two, from the early 1970s New Zealand teachers learned to use individualised assessment strategies such as running records and miscue analysis, promoted following research by Clay (1979), and process writing, following Graves (1983). Teacher education texts of the 1960s (and beyond) were replete with descriptions of how to educate children "naturally". Although she

trained at Auckland Teachers' College in the early 1970s, Anne's teacher education curriculum would have been similar to that described by Harold (in Middleton & May, 1997) as leaving you with the "feeling that the child was the epitome and everything you did was centred on the child" (p. 273). As Walkerdine explains

What was proposed was a process – a scientific process – whereby the schoolroom could become a laboratory, where development could be watched, monitored and set along the right path. (Walkerdine, 1992, pp. 17-18)

This progressive pedagogy, with the focus on the child as a developing person (McInerney & McInerney, 1998), conceptualised the teacher-as-scientist (Walkerdine, 1984). Anne herself stated that "it was all Piaget while I was at Auckland; however, Piaget leaves lots of gaps for me now".

#### Using Records to Increase the Efficiency of Formative Assessment

Anne, however, had combined her child-centred pedagogy with the discourses current in the national curriculum documents, often using the achievement objectives as the context within which her teaching took place. The day I began my note taking and audiotaping, Anne and Lisa had organised their teaching within the context of cooking and baking. This lesson was planned to meet achievement objectives from the science curriculum, and Anne's group contained 14 children who were involved in making queen cakes. Anne had the recipe laminated on card and used it as a shared reading exercise.<sup>27</sup> As six of the children had only been at school for about two weeks, Anne commented (in an aside to me) that she was interested that one of them, a boy, pointed out the word "mix" to her and indicated that he had remembered it.

The next activity included counting the paper cups in which the cakes would be baked. The recipe stated that 22 would be necessary and Anne noticed

---

<sup>27</sup> Shared reading involves the teacher and a group of students in reading a script together, usually with enlarged text so that it can be seen easily by all members of the group.

that one child read this as 21. Anne asked again and another child decided that the numeral was 22. Although the children counted the cup cake cups in unison as Anne placed them on the table, she noticed who was responding to her questions about what would be one more than the amount as she placed each one. She repeated this counting process as the cups were filled with the mixture and noticed that one child within the group could count right through to 22 consistently.

The cakes were counted again as they were placed on the baking tray and Anne asked the children about the hot oven. She noted (and commented to me later) that Carla, who had answered about the oven being “hot”, was a child who usually did not take part in classroom discussions or interactions and that this was an important contribution in terms of her communication skills.

While the cakes were cooking, Anne capitalised on the context and produced a copy of the reading book *Who's Going to Lick the Bowl?* (Melser, 1981) for each of the children in the group. After introducing the book, Anne listened and watched carefully as the children “read” it. During this the bell for morning interval rang and Anne dismissed the children, helped some to find their morning tea and checked that Carla knew what was happening and that she was expected to go outside to play. Anne then picked up her recording book and recorded information against the names of several children. For example, she noted the child who could count confidently to 22, the fact that Carla was responding appropriately to classroom questions, the children who had easily read the reading book she had used following the baking, as well as who had demonstrated the science understandings that were the “planned” objectives of the lesson.

Although this was a science lesson, Anne noticed and acted upon specific instances of competence with numeracy, oral language and literacy, as well as the children’s explanations of how the cakes changed from the individual ingredients through mixing and baking to the final result as queen cakes. This noticing and acting upon significant learning happened throughout the day. Although the teaching was planned according to the achievement objectives from particular curriculum areas, Anne noticed and acted upon information about individual

students' learning across the curriculum but mostly as it related to the development of social, oracy, literacy and mathematics development. In this way she managed to build up an emerging picture of how each child within her class was developing and learning across the curriculum. Anne's dominant mode of assessing fitted with Bell and Cowie's (1997) description of interactive formative assessment rather than the more formal planned formative assessment.

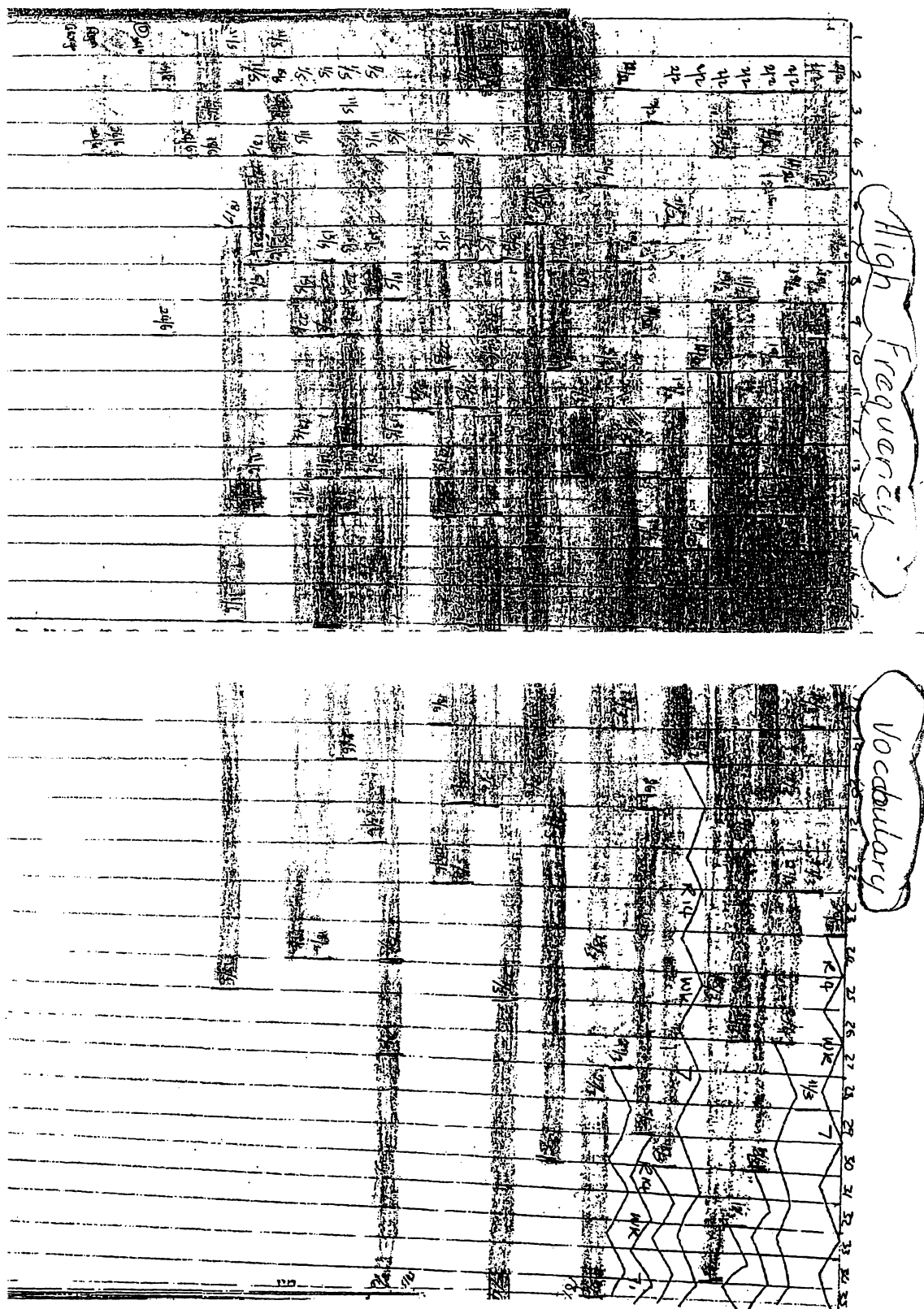
Anne had designed a way of systematically recording significant information about each child in these areas (see Figures 8.8 – 8.12 for some examples). She explained her use of her data book to me in this way

So it's with me, nice and close to me whenever I want it. Participation and everything. And notes, and I find now, that because I know the document [her data book] so well...that as the children are doing their work, I can quickly just note down in it what's happening or what they need. I have it on hand. I couldn't work without it. I need to refer back every so often.

Anne not only recorded to keep track of progress, she used her recorded information. She said she couldn't work without it and that she needed to refer back to it every so often. Although she dated her recording infrequently, I observed how she checked to see if significant behaviour she was noticing was actually "new" for the child concerned. Rather than use a code against objectives, Anne checked off each ability as it was achieved and kept a detailed record of several aspects of literacy development (see, for example, Figures 8.10, 8.11 and 8.12). I saw her call a group together towards the end of a session in which she had been roving and recording as she taught. Sending the rest of the class to work with the second teacher in the room, she focused specifically on teaching the group to use the beginning sounds of words when they were not sure how to proceed in reading. This only lasted for a few minutes, but demonstrated how Anne used her records to calibrate her teaching.





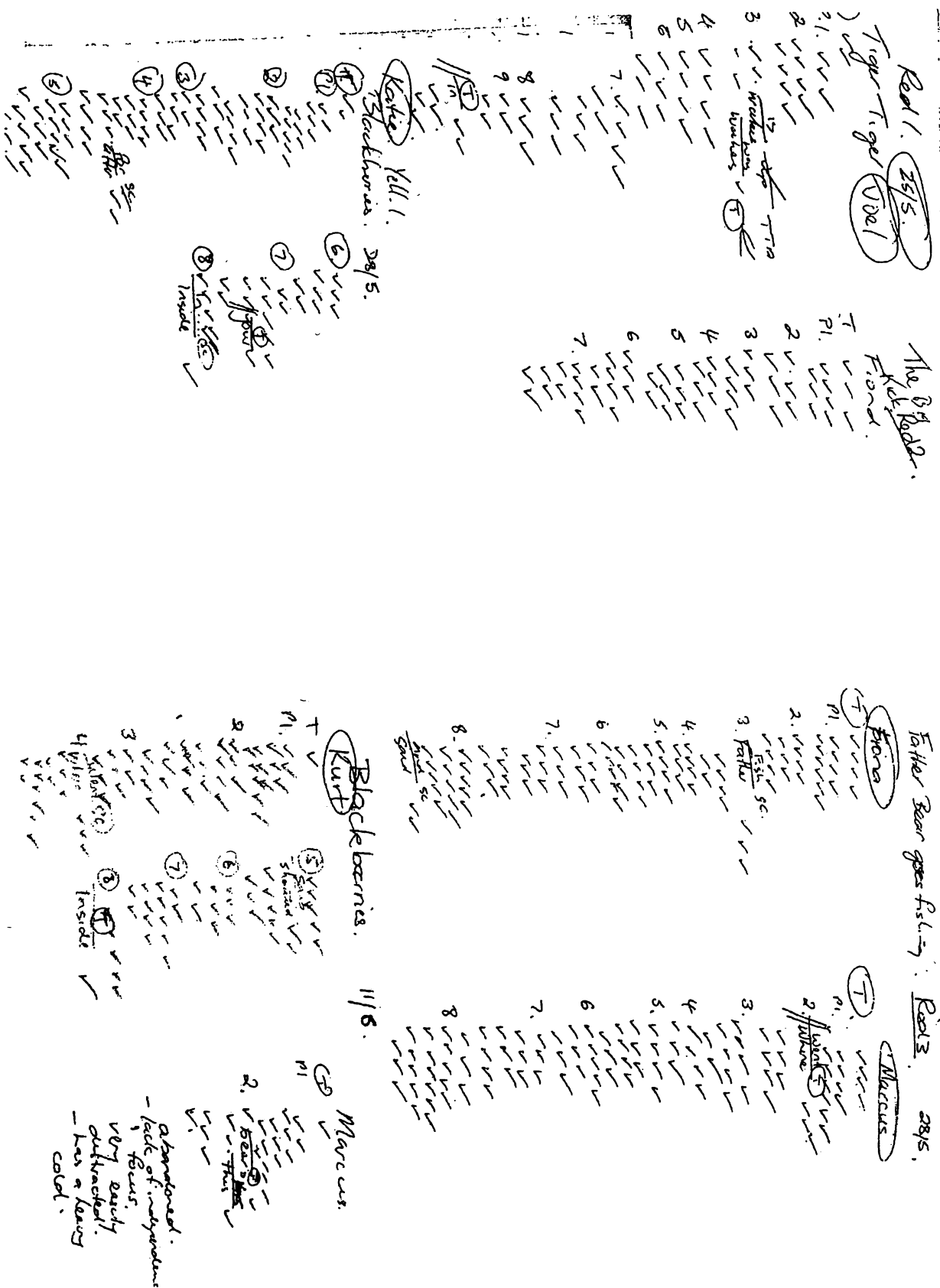


**Figure 8.10** Page from Anne's data book showing how she kept track of high frequency words in reading vocabulary



|  |  |                                                                                                                   |
|--|--|-------------------------------------------------------------------------------------------------------------------|
|  |  | Written 1                                                                                                         |
|  |  | Reading Skills                                                                                                    |
|  |  | Emergent                                                                                                          |
|  |  | <ul style="list-style-type: none"> <li>• Front of book</li> </ul>                                                 |
|  |  | <ul style="list-style-type: none"> <li>• Print contains a message</li> </ul>                                      |
|  |  | <ul style="list-style-type: none"> <li>• Where to start</li> </ul>                                                |
|  |  | <ul style="list-style-type: none"> <li>• Left to right</li> </ul>                                                 |
|  |  | <ul style="list-style-type: none"> <li>• Return sweep</li> </ul>                                                  |
|  |  | <ul style="list-style-type: none"> <li>• Word by word matching</li> </ul>                                         |
|  |  | <ul style="list-style-type: none"> <li>• First and last</li> </ul>                                                |
|  |  | <ul style="list-style-type: none"> <li>• Print is up the right way</li> </ul>                                     |
|  |  | <ul style="list-style-type: none"> <li>• Left page before right</li> </ul>                                        |
|  |  | <ul style="list-style-type: none"> <li>• Full stop</li> </ul>                                                     |
|  |  | <ul style="list-style-type: none"> <li>• Question mark</li> </ul>                                                 |
|  |  | <ul style="list-style-type: none"> <li>• Quotation marks</li> </ul>                                               |
|  |  | <ul style="list-style-type: none"> <li>• Capital &amp; lower case letter correspondence</li> </ul>                |
|  |  | <ul style="list-style-type: none"> <li>• Letters/words - recognises the difference and knows the terms</li> </ul> |
|  |  | <ul style="list-style-type: none"> <li>• First and last letters of words</li> </ul>                               |
|  |  | <ul style="list-style-type: none"> <li>• Capital letters</li> </ul>                                               |
|  |  | Comments                                                                                                          |
|  |  | Red 1                                                                                                             |
|  |  | <ul style="list-style-type: none"> <li>• Full stop</li> </ul>                                                     |
|  |  | <ul style="list-style-type: none"> <li>• Comma</li> </ul>                                                         |
|  |  | <ul style="list-style-type: none"> <li>• Question mark</li> </ul>                                                 |
|  |  | <ul style="list-style-type: none"> <li>• Picture discussion</li> </ul>                                            |
|  |  | <ul style="list-style-type: none"> <li>• Initial sounds of unknown words</li> </ul>                               |
|  |  | <ul style="list-style-type: none"> <li>• Beginning to read for meaning</li> </ul>                                 |
|  |  | Comments                                                                                                          |

**Figure 8.11** Page from Anne's data book showing how she kept track of achievement in concepts about print



**Figure 8.12** Pages from Anne's data book showing how she recorded running records on a daily basis

Although Anne's "data book" was central to her ability to keep track of each individual, she also kept other records about their learning. For example, star-shaped cards were paper clipped into the children's own exercise books to remind both teacher and child about recent progress. One such card clipped into Sam's written expression book read "I am really clever. I don't need a picture any more. I write, write, write." This allowed the teacher to remember, as she worked, what to focus on in writing for each child and to reinforce or extend this aspect with the child as the teaching process proceeded. It was also there later, should she want to record in her data book, or share progress with a parent or another teacher. Anne told me that because the "star cards" were different colours, the children looked forward to the days that they got a "new colour". "They learn that they are making progress for themselves with that new colour. "

A third way in which Anne recorded information about learning as she taught was on the clipboards each child kept among their personal possessions. These were used as a place to collect notices for home and where Anne could record information for parents/caregivers when necessary. Thus accountability to parents was built in as part of the programme rather than being an extra to it. I observed her moving around the tables, assisting children to practise the handwriting examples on the bottom of the daily news sheet. She gave feedback to each of the children as she went and jotted notes on some clipboards for parents about which letter formations needed to be practised at home.

Even though this description is restricted to small parts of a morning's work in Anne's classroom, the complexity of the assessment activity is apparent. The science lesson was planned to teach and check up on a few selected science objectives, and although this was achieved, the assessment behaviour included many more aspects of learning than could be formally planned for. Anne's ability to hold all the progressions in her head, elicit responses, notice important learning signposts, respond to and act upon these when they occurred and record them when she believed they were significant for each child, was stunning. All this took place at a relatively fast pace. The systems Anne had set up for recording and

reporting within her classroom action did not slow the teaching down or take her away from the teaching situation.

### Shared Teaching: Implications for Assessment

The shared teaching situation also allowed Anne and Lisa to attend to more formal diagnostic assessment without interruptions. For example, while Lisa was calling the roll and beginning the day, I observed Anne undertaking a running record of reading behaviour with one child in a quiet corner. Anne recorded the reading, asked some comprehension questions and checked knowledge of compound words from the story. She recorded anecdotal notes about the event in her data book and popped the recording sheet into the child's individual folder for later analysis.

The shared situation also facilitated discussion amongst the teachers about particular children's progress. I observed Anne and Jane (the part-time teacher) reading and discussing a description of the baking lesson written by one of the older children. They read the story, leafed through his book to note his progress and discussed what they might encourage him to do next to improve his writing. They made some decisions about assisting him to improve his spelling and encouraging him to continue to write at length. Although this discussion was not recorded, it was important in meeting the learning needs of that particular child and also in continuing to flesh out, in the teachers' minds, what specific learning/teaching steps were appropriate. Nias et al. (cited in Woods et al., 1997) identified this as professional development within the workplace with others on a daily basis. They suggested that many mid-career teachers actively search for professional extension, viewing their colleagues as one means of achieving this. These studies revealed that these teachers found successful peer learning rewarding and exciting. This appeared to be the case with Anne. Not only did her colleagues provide a forum for discussing children's progress for the children's sake, but Anne, Jane and Lisa continually modified the curriculum through this type of interaction.

A further effect of Anne's approach to assessment was the impact it made on the teachers working with her. Lisa, a second-year teacher, had been working in the same classroom for most of the year. From the interviews carried out with her the year before, I had classified Lisa's approach to assessment as very mixed. By the time I interviewed Lisa for a third time she had been team-teaching with Anne for a few months. The change in Lisa's description of her assessment and teaching was significant. She had modelled her approach on Anne's, using many of the outcomes Anne used in tracking literacy and numeracy development. In fact, because Anne had written a set of learning outcomes for junior numeracy development, many of the teachers in the junior department of Arawhata School were doing the same. But Lisa explained that she and Anne took turns at planning new units of work and that this included deciding what learning outcomes they would add to their data books to assess. It appeared that such an integrated systematic style of assessing was not dependent on the teacher being very experienced. The fact that Lisa had been able to learn and operate the system in only her second year of teaching indicated that such practices could be learned, particularly under the surveillance and guidance of an experienced teacher like Anne.

### Theorising Anne's Practice

Much of this description of Anne's assessment practice fits within that described by Bell and Cowie (1997) and Cowie (1997) as formative assessment. Brown (1996) made the distinction between "strong" and "weak" formative assessment, describing the strong variety as taking into account how an individual makes sense of the world and suggesting that this is usually spontaneous and the information gained ephemeral and acted upon rather than recorded. She described weak formative assessment as being more like repeated summative assessment, in that it is tied to relatively fixed criteria, can be pre-planned and results in relatively crude adjustments to teaching and learning for the whole class. But Anne's assessment practice also included monitoring progress against relatively fixed criteria, recording significant progress and reporting elements of learning, at times,

simultaneously. In this sense, Anne was not only working on improving children's learning across the curriculum as she taught, but she was using assessment for both weak and strong formative purposes, diagnostically and summatively at the same time.

In Bernstein's (1971) terms, Anne was using an integrated approach where the boundaries between the individual subjects (as they are found in the curriculum documents) are blurred (p. 53). At the same time, she had developed a systematic way of accounting for learning in each curriculum area by recording significant progress under the appropriate curriculum area heading. Anne's way of blending child-centred formative assessment with the requirement to monitor progress against the achievement objectives was described in Chapter Seven as an "integrated systematic" style of classroom assessment. The observations in Anne's classroom reinforced the interview data that indicated some teachers had developed quite complex systems of monitoring and informing the learning of individual students.

As reviewed in Chapter Four, Crooks (1988) and Black and Wiliam (1998a) have conducted extensive investigations of the research literature and both reached the conclusion that formative assessment does improve learning. Crooks found that feedback to students should focus on the task, should be given regularly and while still relevant and should be specific to the task. "Most vital of all the messages emerging from this review" (Crooks, 1988, p. 470) was that the assessments must emphasise the skills, knowledge and attitudes perceived to be most important, however difficult the technical problems that this might cause. In sorting out for herself what she felt were the most important skills, knowledge and attitudes for learning in each of the curriculum areas and continually focusing on these, Anne was constantly giving quality feedback and monitoring progress simultaneously.

The literature reviewed by Black and Wiliam demonstrated that there is evidence that formative assessment can "produce significant, and often substantial, learning gains" (1998b, p. 3). The reports they studied featured enhanced formative assessment as a common feature and suggested that enhancing feedback

between those taught and the teacher, actively involving the students in learning, adjusting teaching during teaching and engaging students in self-assessment in ways which positively affect motivation and self-esteem were all important factors (p. 3). Even though the children Anne was teaching were 5- and 6-year-olds, she had developed sophisticated ways of achieving all of these characteristics in her teaching. The children were actively involved in well-paced classroom activities that required their interaction and attention. Anne used devices such as star cards both to record progress and remind the children about their current goals and achievements. Her constant interaction and feedback with the children kept them focused on self-assessing their learning. For example, the star card clipped into Sam's written expression book that read "I am really clever. I don't need a picture any more. I write, write, write" allowed the teacher to remember, as she worked, what to focus on in writing for that child and to reinforce or extend this aspect with him as the teaching process proceeded. It was written with Sam's help in a way that modelled self-assessment and gave voice to Sam as an assessor as well as a learner. In this way, much of her recording of assessment was placed in the student's work, not just in separate records for the ERO and others.

### Disciplining Learning – and Teaching

Anne's classroom practice can also be seen to be productive from a Foucauldian perspective. In focusing her gaze on what she counted as the most important knowledge, skills and attitudes Anne was able to refine her assessment technology. She skilfully combined her powers of observation with her expectations of learners in an ongoing process of examination (Foucault, 1977). For example, over the years Anne had clarified for herself exactly what children in her class needed to know and be able to do in becoming literate and numerate. She focused on these aspects constantly during teaching as she observed every learner in action. She recorded their progress against these indicators as they were learning and provided continual feedback on what it was they were trying to achieve and how they were progressing. Furthermore, she involved each child in assessing how well he or she was progressing in each aspect. Foucault stated that

the aim of his work had been to “create a history of the different modes by which, in our culture, human beings are made subjects” (1982, p. 208). In rejecting an *a priori* theory of the subject, Foucault defined the basic question as how the subject constitutes him or herself through practices that are basically related to power and knowledge. Self-assessment, where children come to see themselves as knowing and skilful people, increases the effect of the teacher’s assessment. Children, as learners, come to take responsibility for their own learning by keeping themselves under surveillance and measuring themselves constantly against the normative expectations that the teacher sets up. As argued in Chapter Seven, it is this self-examination that disciplines learning so efficiently.

### **Summary and Comment**

The focus of this chapter was specifically upon the three teachers, Rose, Luke and Anne, in order to elucidate the effects the multiple and competing discourses of assessment were having on their practices. Although it was possible to recognise these teachers as exemplifying each of the three approaches to assessment described in Chapter Seven as unit, head note and integrated systematic assessment, the portraits in this chapter also describe how personal professional knowledge accumulates and influences individual teachers in their practice. In each of these examples, the teachers monitored their students, the students monitored themselves and each other, and the teachers monitored each other and themselves. “Surveillance singles out individuals, regulates behaviour, and enables comparisons to be made” (Gore, 1998, p. 236).

In contrast to those who view Foucault as a theorist who “wins by building an increasingly closed and terrifying machine” (Jameson, cited in Blacker, 1998), the analysis of these three teachers’ assessment practices enabled me to understand the productiveness of disciplinary power (Foucault, 1977). For example, in these portraits I have demonstrated both the productiveness of the examination (assessment) in generating and sustaining learning, as well as the productiveness of examination in regulating teachers’ behaviour in self-managing schools. I



found myself continuously struggling with the conception of power as a disciplinary force, often slipping into the trap of regarding power as sinister. It took much effort and reconsideration to understand how power's pervasiveness was not a "problem" but a "given", and to embrace Foucault's point that "a society without power relations can only be an abstraction" (1982, p. 222). However,

to say that there cannot be a society without power relations is not to say either that those which are established are necessary, or, in any case, that power constitutes a fatality at the heart of societies, such that it cannot be undermined. (Foucault, 1982, p. 223)

The purpose of revealing power at work as assessment in the classrooms I investigated allowed me to reveal the (surveillance) practices that disciplined teachers to concentrate their efforts on recording and reporting, which diverted many from using formative assessment strategies. At the micro-level of formative assessment, power operates to make the individual student in self-managing primary schools as literate, numerate and "useful" in our society. Formative assessment is the mechanism by which teachers move the learning forward through all the techniques of power that Foucault and others (for example, Gore, 1998; Middleton, 1998; 1999) have described as surveillance, normalisation, exclusion, classification, distribution, individualisation, totalisation and regulation. However, due to the weight and power of the accountability discourses, introduced in Chapter Three, teachers using a unit assessment approach were disciplined into using more summative approaches to assessment than head noters, who resisted much of the checklisting, or those using an integrated systematic approach, who defined their own assessment categories and only recorded significant, useful information. While summative checklisting enables some analysis of learning needs and outcomes, its main purpose is accountability, so that proximity to the curriculum norm may be measured and made visible.

Moreover, measuring individuals, classes and whole school populations in relation to the stated curriculum norm, as well as diverting teachers from raising achievement through more formative assessment, has the effect of allowing others

– namely, principals, boards of trustees and the ERO – to observe and monitor teachers, as well as students, more closely. As Foucault explains

Any panoptic institution...may without difficulty be subjected to such irregular and constant inspections: and not only by the appointed inspectors, but also by the public; any member of society will have the right to come and see with his own eyes how the schools, hospitals, factories, prisons function.... This panopticon, subtly arranged so that an observer may observe, at a glance, so many different individuals, also enables everyone to come and observe any of the observers. The seeing machine was once a sort of dark room into which individuals spied; it has become a transparent building in which the exercise of power may be supervised by society as a whole. (1977, p. 207)

Rather than regarding this as a sinister finding, another interpretation is that, knowing formative assessment promotes learning, the challenge is to find ways of checking up on learning in schools that do not discipline teachers away from predominantly formative practices. Major clues to how this could be achieved are to be found in the practices of formative assessment examined in this thesis, and by investigating the ways in which Luke and Anne met the accountability demands placed upon them, ways to reduce the burden of summative recording and reporting have been revealed.

Although the focus of the research reported here has been on the classroom, the findings demonstrate that it is the entire context of self-managing schools, along with national and school policies of assessment and school review, that has disciplined teachers' practice towards more summative or formative approaches. Chapter Nine discusses the findings reported in Chapters Six to Eight and explores the implications of this investigation for raising achievement in primary classrooms, for changing the emphasis in teacher education about assessment of learning, and for policy making and new assessment initiatives. It also addresses the limitations of the research reported here and makes recommendations for further research.

## **CHAPTER NINE**

### **LESSONS FROM THE CLASSROOM: REVISIONING THE ASSESSMENT LANDSCAPE**

If schools are not to become “museums of modernity” we urgently need to rediscover intuition and engage with the challenge it presents for education. Only by doing so can we exchange the discipline of the ‘assessment society’ for the liberation of the learning community. (Broadfoot, 2000, p. 216)

My role – and that is too emphatic a word – is to show people that they are much freer than they feel, that people accept as truth, as evidence, some themes which have been built up at a certain moment in history, and that this so-called evidence can be criticized and destroyed. (Foucault, cited in Martin, 1988, p. 10)

This participative case study took as its starting point a current “problem”, namely, teachers’ concerns about assessment in self-managing schools. Prompted by teachers’ stress and confusion about their assessment role, I employed Foucault’s theoretical notions of discourse, disciplinary power/knowledge and the examination as a springboard to understanding what the problem might be. While it is certainly possible to study the effects on teachers without recourse to Foucault’s interpretive analytics, re-reading contemporary educational and managerialist discourses of assessment in the light of Foucault’s work had the advantage of disrupting accepted rationalities, taken-for-granted claims and plausibility structures about assessment, and raised a number of questions about the discursive practices which foreground assessment presently and link it to questions of economy and power (Comber, 1997, p. 390). To do this, it was necessary to “know the historical conditions which motivate our conceptualisation. We need a historical awareness of our present circumstance” (Foucault, 1982, p. 209).

First, therefore, I argued that since education became compulsory in New Zealand, rather than a smooth process of improvement portrayed by the dominant

liberal humanist discourses of Western societies, there have been ruptures and breaks in the education system and society as a whole that have promoted various dominant assessment discourses for teachers' attention. Furthermore, although the more powerful discourses held sway at any one time, the traces of other competing discourses of assessment and accountability were uncovered in teachers' practices and the assessment literature. Moving to the recent past, I argued that Tomorrow's Schools, which brought about administrative and curriculum changes, caused a major rupture in teachers' assessment discourses in the early 1990s. In summary, these changes caused a swing to the promotion of competition between schools and the use of managerial accountability practices to accommodate this marketisation. Concurrently with, but in contrast to, these national policy changes for schools, teachers were also impacted by professional discourses of assessment and learning that generally promoted formative over summative accountability functions of assessment (for example, Bell & Cowie, 1997; Black & Wiliam, 1998b; Broadfoot, 2000; Sadler, 1989).

The analysis of the data from the empirical fieldwork in this thesis demonstrated that the discursive shift from education as a human right to education as a government investment (Comber, 1997; Fitzsimons, 1996) that took place in New Zealand in the last decade has impacted on teachers' assessment practices. The evidence reported in Chapters Six, Seven and Eight reveals that primary teachers in self-managing schools have been subjected to increased managerial surveillance through the mechanisms of school and syndicate teamwork, appraisal, performance management and external audit. Due to this increase in managerialism teachers have put themselves under pressure to account for their own practice. They have done this by finding ways to meet the documentation demands of the system, and in the process, have produced a summative shift in their assessment practice.

But the multiple and often contradictory subject positions exemplified in the narrative portraits in Chapter Eight preclude a single interpretation of the ways in which school self-management and curriculum changes have affected teachers, specifically in assessment. As Kvale (1996) and Pagano (1990, cited in Munro,

1998) remind us, there is more than one way to tell a story and there is more than one story. A single interpretation of these teachers' stories is not only impossible, but also undesirable. I can in no way provide some final interpretation of these stories that will enlighten policy makers, researchers, teacher educators or practitioners about how to improve assessment and enhance learning. And, in fact, that is not my purpose. How to reach some sort of conclusions from the data presented in this report is the "contradiction with which I engage" (Munro, 1998).

Rather than theorise new models of assessment practice, my intention is to engage in a discussion that questions unitary readings of concepts like accountability, monitoring, recording, reporting, progress and achievement. Because teachers' practices in self-managing schools are both constituted by and constitutive of the multiple and competing discourses of assessment, their understanding and implementation of the different aspects of assessment vary. They vary both within and between classrooms. Within the same classroom, accountability can mean the teacher's accountability to students and their parents, colleagues, the management team, the board of trustees or taxpayers in general. Different teachers within the same school might also interpret what is required of them in terms of accountability differently. The same is so in terms of how each teacher understands and interprets other aspects of assessment, such as monitoring, progress and even achievement.

Furthermore, although teachers in this study were observed using the kinds of formative assessment strategies known to improve learning and raise standards, most did not identify these strategies as assessment and often understood the use of continuous summative assessment as formative. This is an important finding for teacher education. The portraits in Chapter Eight are illustrative of how multiple competing assessment discourses, among others, have influenced the classroom practices of three teachers. This chapter uses the lessons from the schools and classrooms in which the research was carried out to establish how summative assessment has come to have such a dominant role in the 1990s and to suggest how and why formative assessment might be promoted within self-

managing schools. The impacts of this research on practice are described, and, finally, the limitations of the study and areas for further research are addressed.

### **A Summative Turn in Response to Accountability Demands**

As introduced above, one reading of the stories told in this thesis is that managerial discourses of accountability have produced teacher practices that have diverted attention away from assessment for formative purposes. This finding is supported by evidence from other investigations into teachers' assessment in New Zealand during this period (Dixon, 1999; J. Hill, 1998a; Wylie, 1994; 1998) and by other international evidence (for example, Gipps et al., 1995; Hargreaves & Evans, 1997; Torrance & Pryor, 1995; Woods et al., 1997). The analysis of teachers' practices derived from the interview responses in both the case study schools and the teachers interviewed nationally indicated that there were some typical ways adopted by schools, and the teachers within them, to meet these demands. As argued in Chapter Seven, these approaches provide varied amounts of space within which teachers could move to implement formative assessment strategies. It is of concern that, although there is strong evidence to suggest that formative assessment can raise standards (Black & Wiliam, 1998b) and that such a strong case cannot be made for summative assessment or for using the analysis of school-wide standards and trends, so much effort and pressure is being exerted nationally to measure and report outcomes. In responding to external audit and Ministry of Education demands to measure achievement against the achievement objectives across the curriculum, and through designing and disciplining themselves to fill in a proliferation of checklists to achieve this, teachers in primary schools in New Zealand have reduced their commitment to carry out formative assessment.

The Education Review Office (1999) report on the assessment of student learning also provided support for this finding. It indicated that the analysis of the data it gathered about teachers' assessment practices from 118 primary schools in

late 1998 revealed teachers were collecting much information for managerial purposes.

Many teachers have attempted to meet [assessment] requirements without first giving careful consideration to questions of purpose. As a result they gather a great deal of information that is not useful to them and is not used by them. They also gather information without consideration of the quality of the data they record. This leads to the use of information, especially for summative purposes, which may not be appropriate for that purpose. (Education Review Office, 1999, p. 3)

Ironically, evidence reported in this thesis suggested that these teachers were gathering the type of summative information, labelled by the ERO as inappropriate, *for* the ERO so that they could be seen to be meeting the requirements of the NEG's and the NAG's. Teachers in this study, throughout New Zealand, explained that they "have to" do these types of recording in order to be accountable. Rose told me that since the restructuring of education "everyone's far more accountable, right through, all the way down the line...not just in education, it's in all areas now, everyone's more accountable". Looking through Foucauldian lenses, Rose is expressing the prevailing and powerful discourses of accountability that have come to dominate teaching since the early 1990s. In this respect the term "accountability" has been used by governments in New Zealand to establish a discursive consensus which constructs schools and teachers as being in need of external regulation (Fitzsimons, 1996; Poulson, 1996).

Accountability may be seen less as a moral obligation to clients or colleagues, either individually or collectively, than as an aspect of the disciplinary technology (Foucault, 1977) by which the work of teachers and schools is surveyed and controlled. (Poulson, 1996, p. 585)

Timperley and Robinson (1997) argue that when introducing radical reforms to educational administration in 1989, the New Zealand government was committed to making the educational sector "more accountable and more responsive to stake holders" (p. 157). In addition to the regular visits of the ERO to check that the NEG's and the NAG's (amongst other things) had been

operationalised, appraisal and performance management systems became the mandated responsibility of the boards of trustees, the governance body of each school (Timperley & Robinson, 1997). Willis (1994) has argued that a performance model of assessment such as this does not necessarily facilitate high-quality learning. Refining and extending Blackmore's (1988) description, Willis argued that using learning outcomes for measuring progress is aligned with a technicist approach and is "a feature of market and management accountability" (1994, p. 171). Codd (1998), too, pointed out that in a number of countries, including Britain, North America, Australia and New Zealand, there have been moves towards decentralisation of specific administrative functions with an emphasis on more managerial control and accountability at the school level. Essentially the model of management has been an industrial one and the role of the principal has been redefined in terms similar to that of the executive or manager. Teachers are increasingly regarded as workers rather than professionals, with their productivity assessed in terms of the test results and examination performances of their students (see also Broadfoot, 2000).

Another reading of the stories told in this thesis is that immersing teachers within an outcomes-based curriculum has caused many to produce a technicist approach to their teaching because for them it has seemed the most manageable way of meeting the competing demands on their time. Required to monitor against the achievement objectives (Ministry of Education, 1993b) and to carry out school and self-improvement (H. Lee & Lee, 1998), schools nationally have attempted to gather data on as many achievement objectives or learning outcomes as possible. The evidence from at least one school in this thesis (see Chapter Six) was that the ERO required schools to monitor against all achievement objectives in all curriculum areas. The advice from the ERO at the time the field work for this thesis took place, documented in Chapter Six, was for schools to aggregate school-wide results to look for patterns that might allow them to see how well students in their school were performing (Education Review Office, 1995a; 1995c). To achieve this, some schools have begun to use software packages that aggregate information about how children have met the outcomes expected of them so that



teachers can analyse the results for patterns and trends. While the results can provide useful information to assist with future planning for programmes and resources, gathering this sort of data does not usually assist teachers to modify their teaching in classrooms. In fact, when teachers are required to gather large amounts of achievement data for these software systems, the danger is that it will take time away from more valuable formative assessment and planning opportunities in the classroom.

### Multiple Stories of Assessment

The unit assessment approach described in this thesis, which involved defining clear objectives, undertaking performance assessment and recording progress against the defined objectives, can be seen to fit neatly within the discourses of economic rationalism, managerialism and accountability. Both Rose and Luke used record systems consistent with convergent assessment (Pryor & Torrance, 1996; Torrance & Pryor, 1995), which they regarded as formative assessment. Evocative of Bernstein's "collection type" curriculum, "where the learner has to collect a group of favoured contents in order to satisfy some criteria of evaluation" (Bernstein, 1971, p. 49), these systems were productive of learners as educated in terms of the dominant discourses contained in the national curriculum.

But, as has been described in Chapter Eight, teachers using a unit assessment approach, while outwardly complying with this contractual compliance form of accountability (Codd, 1999), concurrently used formative strategies that they had intuitively built up over their teaching experience. Another way to describe this is that these teachers used their personal practical knowledge of learning, teaching and assessment (Clandinin, 1998; Clandinin & Connelly, 1995) to make decisions about teaching in action.

What we mean by teachers' [personal practical] knowledge is that a body of convictions and meanings, conscious or unconscious, that have arisen from experience (intimate, social, traditional) and that are expressed in a person's practices....

When we see practice, we see personal practical knowledge at work. (Clandinin & Connelly, 1995, p. 7)

So, while most of the teachers who participated in this investigation thought of assessment as the documentation they carried out to monitor students' learning against the achievement objectives, they also constantly used their personal mental progressions of learning in the various areas of learning to respond formatively to learners. Both Rose and Luke, through working in self-managing teams, had developed mastery-type codes as a cover story (Clandinin & Connelly, 1995) for their assessment practice while, at the same time, they used their head noting techniques to interact with their students as a way to grow their ideas (Harlen, 1998). The wider sample of teachers interviewed for this thesis and the reactions of other teachers at workshops, conferences and in university courses indicated that these findings resonated with their practices. Thus, within the dominant discourses of standards-based assessment and an outcomes-based curriculum – sacred stories of school in New Zealand in the 1990s (Clandinin & Connelly, 1995) – individual teachers often seemed to be using more than one approach to assessment. On the one hand they recorded achievement of the objectives with a coded system, mainly for other people – a cover story in Clandinin and Connelly's terms – and, on the other, they used their mental maps and timelines to interact formatively with students. With the exception of those teachers termed "integrated systematic" assessors, teachers rarely recorded any of the information from their formative interactions (cf. Bell & Cowie, 1997; M. Carr & Cowie, 1997). Because this approach to assessment was largely private, thought of by teachers as teaching or evaluation rather than assessment and took place in classrooms away from the gaze of others (cf. "secret stories", Clandinin & Connelly, 1995), it is likely to be the most difficult to supervise and change.

As well as revealing how teachers can be disciplined into using particular assessment and documentation practices, the analysis in this thesis also indicates that teachers still create space to focus their gaze on what they believe really matters in learning. The fact that Rose struggled to balance a managerial approach with her belief that the most valuable form of assessment was to find out where the

children “were at” and “where they were going”, and to “go on from there”, indicated that unit assessment did not rule her practice completely. As described, Rose spent much of her day interacting formatively with her students, using her “head-map and time line”, in other words her professional judgement, to make teaching decisions, during teaching and learning (Bell & Cowie, 1997). Rose herself, however, did not identify this behaviour as assessment but called it teaching. Luke also referred to his formative assessment practice not as assessment but as evaluation. These results are consistent with Dixon’s (1999) finding that the teachers she interviewed were not clear about the meaning of “formative assessment”. And, as she points out, this may not be surprising given the way in which the national curriculum statements and the ERO reports blur and confuse the distinction between formative and summative assessment.

Anne, however, met the requirements to document progress against the achievement objectives differently, using formative and summative modes of assessment often simultaneously. While she could demonstrate how her own progressions fitted within the achievement objectives, she had internalised sets of progressions which allowed her to use more divergent forms of assessment (Pryor & Torrance, 1996; Torrance & Pryor, 1995) while still keeping track of individual progress. In Bernstein’s terms she was using an integrated code that incorporated the “subordination of previously insulated subjects to some relational ideas – blurring the boundaries between the subjects” (Bernstein, 1971, p. 53). While some “integrationists” (for example, Beane, 1992; Brodhagen, 1998) argue that integration is done by the learner him- or herself through a curriculum planned with the young people for whom it is intended, rather than by teachers alone, the age of the children Anne was working with often precluded this approach in its entirety. Her approach was “integrated” in that the programme for each student was mediated by the teacher across subject boundaries in an attempt to make the curriculum coherent and progressive for each learner. The students were involved in their own curriculum, however, to the extent that Anne involved them continually in self-assessment that focused their learning. In fact, Anne’s

approach revealed traces of “integration” as described by Ball in 1948 (see Chapter Two, p. 37).

### Improving Learning through Formative Assessment

Although none of the teachers used the terms “formative” or “summative” assessment, almost without exception the interview data demonstrated that formative assessment was intrinsically highly valued. As reported in Chapter Six, teachers from the two case study schools all reported that they assessed because they needed to know where children “were at”, and talked of their teaching being “child-centred” to refer to the ways in which they attempted to observe children and act on what they saw immediately. In addition, teachers interviewed in the national sample of schools explained individualised assessment techniques used during teaching and sometimes referred to their application as “using professional judgement”.

The portraits in Chapter Eight provided clear evidence that all teachers use formative assessment strategies. For example, Anne’s classroom assessment practice is her way of bringing to bear her past experience and her intentions of assisting children to learn in a way that she believes will promote that learning. As was described earlier, Anne valued what she learnt from every assessment event. She did not just look for evidence of learning that was confined to the objectives she had planned to check on, she also elicited particular performances and noticed significant learning as it occurred (Bell & Cowie, 1997; Cowie, 1997). Rather than plan just for whole-class lessons or even for groups within her class, Anne focused her attention on individuals within her class. Through her prior experiences as a teacher she had built up her personal practical knowledge and used it within the classroom and outside the classroom in school contexts.

Rose and Luke also brought their personal/practical knowledge to bear on their teaching. While they didn’t appear to regard their formative assessment strategies as assessment, both explained how they gathered learning information, recognised significant features of students’ activity, and responded appropriately to extend children’s learning (Bell & Cowie, 1997; Harlen, 1998). As explained

earlier, Rose talked of this kind of assessment as teaching while Luke called it evaluation.

Thus, in addition to producing policies that discipline teachers to checklist and use a predominance of summative assessment strategies, teachers still make space to use and promote formative assessment practices. For example, in Anne's classroom, Lisa was supervised by Anne and came to use similar systematically integrated assessment techniques. Anne provided a great deal of professional development for Lisa, encouraging her to use assessment formatively throughout her teaching. Although both of these teachers also used unit assessment strategies for some aspects of the curriculum, they prioritised literacy and mathematics in their classroom and used finely tuned formative strategies to collect information about learning so that it could be used almost immediately to influence their teaching. Thus, Anne's surveillance of Lisa's practice was productive of more formative assessment practices while, at the same time, their practices met the requirements of the school policies.

Assessment research (for example, Black & Wiliam, 1998b; Crooks, 1988) is not alone in supporting the belief that it is formative assessment that will lead to better learning and raised standards. Stoll (1998), in the school improvement literature, argued that in an overarching vision of improvement the focus must be on the centrality of learning. She writes that what must be of prime concern is what goes on between teachers and students in their classrooms, and indicated that one particularly important issue is drawing upon current research to gain a more detailed understanding of the learning (and teaching) process. She said this includes understanding

- what learners bring to their learning. In short, learners do not all learn in the same way,
- what motivates learning,
- what constitutes effective learning and how learning takes place,
- the different locations where the process of learning takes place,
- what the implications of learning are for teaching, including the use of new technologies. (Stoll, 1998, pp. 3-4)

By “new technologies” Stoll implies electronic systems of communication and documentation. But, in addition, the quasi-mastery techniques described earlier in this chapter are technologies that have been invented by teachers to make the accountability requirements manageable in their everyday classrooms. Teachers, and, importantly, principals and others in management positions in schools, need to understand the implications of what “what teachers do” does in terms of learning. Both electronic technologies and coding systems will not in and of themselves improve learning and teaching. In fact, the evidence presented in this thesis has demonstrated that such technologies may have unintended consequences which can drive teachers to ticking instead of teaching. However, the example of Anne’s integrated systematic style of assessment, and Rose and Luke’s formative strategies, which communicate to learners what is valued (Cowie, 1997), demonstrate that when teachers have the purpose for assessment clear in their minds, technologies of surveillance and feedback can be used to maximise student learning.

#### Implications for Professional Development

This finding that teachers often do not associate the term “assessment” with the formative assessment practices that they name “teaching” and “evaluation” (and maybe other things as well, such as, “professional judgement” and “intuition”) has important implications for any attempts to improve assessment in order to influence learning. Research in this area to date clearly demonstrates that formative assessment has the greatest potential to improve learning and standards (Black & Wiliam, 1998b; Crooks, 1988), and that teachers can use formative assessment to achieve better results (for example, Bell & Cowie, 1997; Glover & Thomas, 1999; Torrance & Pryor, 1995). If teachers do not recognise their teaching and evaluation as formative assessment, policy makers, teacher educators, professional developers, the ERO and teachers will be “talking past each other” (Metge & Kinloch, 1978, p. 1). In self-managing schools, many have understood “assessment” only as the practices they engage in to meet accountability requirements. To address this situation, professional development and teacher education programmes need to elicit teachers’ ideas about assessment,

teaching and learning, recognise how these ideas and misconceptions might impact on the professional development proposed, and respond in ways that will assist teachers to focus on identifying and improving their formative assessment practices.

The evidence presented in this thesis is that different teachers understand assessment and their accountability obligations differently. For example, although Anne was working in the same school as Rose, and no doubt felt accountable to her colleagues, she did not mention the school's assessment policy as a constraint. In fact it was Anne who had been called upon to assist the syndicate teams within the school to set up their roll books with learning outcomes. She met with the teams and explained how she had developed her "data" book based on her knowledge of what she expected her children to know and be able to do as they worked with her through their first year to 18 months at school. Neither Anne, nor Lisa, who worked with her, used the achievement objectives directly from the curriculum. Nor did they use a three- or four-point scale as described by Rose and others at that school. The teachers in the other syndicates had decided upon these modifications to Anne's system as they sought to respond to the requirements of the curriculum and the NEGs and the NAGs in a manageable way. Both Anne's and Rose's approaches fell within the guidelines that the staff as a whole had decided upon as their school assessment policy. The different ways that the school-wide policy could be interpreted and implemented provided space for teachers to act more formatively or more summatively, depending on how they controlled themselves in their self-managing teams.

Hargreaves (1994) argues that collaboration and collegiality are pivotal to the orthodoxies of change if one of the "most prominent heresies of educational change is the culture of individualism" (p. 186).

Collaboration and collegiality, it is argued, take teacher development beyond personal, idiosyncratic reflection, or dependence on outside experts, to a point where teachers can learn from each other, sharing and developing their expertise together. Research evidence also suggests that the confidence that comes with collegial sharing and support leads to greater readiness to experiment and take risks, and with it a commitment to continuous improvement among teachers as a recognised part

of their professional obligation. In this sense, collaboration and collegiality are seen as forming vital bridges between school improvement and teacher development. Certainly, those aspects of collaboration and collegiality that take the form of shared decision-making and staff consultation are among the factors which are repeatedly identified as correlating with positive school outcomes in studies of school effectiveness. (Hargreaves, 1994, p. 186)

This is a widely supported international discourse, yet, as Hargreaves goes on to warn, the attractions of collaboration and collegiality should be treated with caution. As with my discussion above, Hargreaves alerts readers to the micropolitics of collegiality and collaboration. Questions such as who guides and controls collaboration and collegiality, and how power operates in collegial settings, are critical to the way in which assessment is understood in teachers' work. His analysis raises the danger of "contrived collegiality" (Hargreaves, 1994, p. 208), in which the cooperative principles of human association among teachers in administratively regulated and predicated systems can discipline self-managing teams to unthought through practices. School effectiveness is reliant on knowledgeable as well as efficient leaders.

### **The Context of Self-managing Schools**

The evidence from the teachers that I interviewed in the study indicated that the NEG's and the NAG's, the ERO reports, and the schools' policies were all very powerful influences on teaching and assessment practice. In effect these discursive practices were instrumental in disciplining teachers to produce assessment technologies that described children's learning through the quasi-mastery codes outlined both in this report and by the ERO (1999). As described at Arawhata School, surveillance comprised hierarchical observation by colleagues, and self-surveillance by the teachers, to check that they were keeping track of sufficient achievement objectives. In addition, the teachers needed to comply with school policies on assessment that ensured data were gathered for annual reports to the board of trustees and for the three yearly checks by the ERO.



This evidence is consistent with international findings, particularly the impact of intrusive “official” inspection by the Office for Standards in Education (Ofsted) on the deprofessionalisation of teachers in England and Wales (Hargreaves & Evans, 1997; Woods et al., 1997). In the Woods et al. study, the teachers who were subjected to inspection experienced stress and a range of emotional responses that in turn caused many of them to become “compliant” rather than “enhancing”.

These teachers do not relish managerial opportunities and changes in role, but they accommodate, concur and allow changes to impinge upon them. This was the largest category among our teachers, but we identified wide variation among them, from “supporting”, to “surviving”, to “disturbed” conformists. (Woods et al., 1997, p. 60)

Hargreaves and Evans (1997) reported that Ofsted have disputed such evidence, arguing that 83.5 per cent of head teachers surveyed were satisfied by their experience of inspection. As Hargreaves and Evans point out, though, this argument rests curiously “on data drawn from managers and not the teachers who work for them” (p. 6), and upon administrative surveys which cannot convey the words of teachers that qualitative research “can help to bring to the printed page” (p. 6).

As in Woods et al. (1997) study, the data from the interviews and observations in the investigation reported in this thesis demonstrate that teachers’ practices are not uniform. They vary from teacher to teacher and from school to school. While summative assessment strategies dominated practice in many cases, this was not an exclusive picture. Some teachers were able to prioritise formative assessment in their teaching by incorporating the requirements to account within their existing assessment strategies, while others resisted the requirement to record against the achievement objectives until a suitable time presented itself.

Not only did these mechanisms ensure compliance but they also produced strong emotional responses from the teachers. For example, Sam stated that she felt guilty if she didn’t document everything she observed the children achieving (see Chapter Six). Clandinin and Connelly (1995) emphasise the importance of

everyday life, personal history, social history and the moral role of other teachers in determining teachers' practices and emotional responses. Likewise, Hargreaves (1997) stresses how important teachers' emotions are in the process of educational change. Because teachers make heavy emotional investments in the whole web of significant and meaningful relationships that make up the work of schools, ignoring emotional responses can lead to disastrous consequences. In Sam's case she had left teaching as a result.

Teachers in this research often talked about assessment as a battle, struggling between what they perceived they should be doing for ERO and what they believed assisted learning. For example, Luke explained that he had learned that assessment was measuring against objectives and recording outcomes, whereas evaluation was in his mind and took learners on to new learning. He had resolved this battle for the meantime by using a coding system to record progress against the achievement objectives from time to time, especially before ERO or anyone in a senior position was due to check up on him. He explained that he had learned this coding system through syndicate and staff meetings following an ERO review at the first school he taught at.

### Concertive Control: School Policy and Teamwork as Normalisation Devices

To a large extent, school and syndicate policies rather than bureaucratic rules have become a major factor regulating teachers' practice in self-managing schools. Influenced by national requirements (the NEGs and the NAGs and the curriculum documents), the ERO reports that they received, and the current imperative for teachers to be accountable to communities and the public at large, schools set up these policies. As self-managing teams, schools and the syndicates within them determined for themselves the rules and norms that teachers would abide by in their day-to-day work, such as taking a running record for each student at least once a term and recording achievement in roll books with an agreed code. The team themselves enforce these norms. Teachers in many of the interviews referred to the assessment policies that guided their practice and explained, like Rose, that during appraisals other team members checked up on data collection,

whether the minimum running records had been done and whether the prescribed samples of work had been collected.

In small schools, such as Waimaka, the decision-making team is usually the staff as a whole. In larger schools, like Arawhata, whole-school policies tend to be implemented through agreed practices in the teams (syndicates, development groups). Self-managing teams regulate their own behaviour through “concertive control” (Barker, 1993, p. 408). Contrary to the notion that self-management frees schools and teams from the “iron cage” of bureaucracy (Weber, 1978, cited in Barker, 1993), the evidence here is that it constrains an organisation’s members even more powerfully. Barker (1993), investigating self-managing teams in business organizations, explained that this more effective system of control comes from the authority and power team-mates exercise on each other as peer managers. He argues that this peer power increases the total amount of control through two important dynamics.

The first is that concertive workers have created this system through their own shared value consensus, which they enforce on each other.... They have put themselves under their own eye of the norm, resulting in a powerful system of control.

The second reason for the increased power of concertive control is the way that it becomes manifest is less apparent than bureaucratic control. Team members are relatively unaware of how the system they created actually controls their actions. Concertive control is much more subtle than a supervisor telling a group of workers what to do... the workers create a value-based system of control and then invest themselves in it through their strong identification with the system. Because of this identification, the team members are socially constructed by the system they have created. When this happens, the team members readily accept that they are controlling their own actions. It seems natural, and they willingly submit to their own control systems. (Barker, 1993, p. 434)

Barker continues by arguing that his analysis suggests that rather than freeing workers from the iron cage of bureaucratic controls, an ironic paradox occurs in which self-management causes the iron cage to become stronger. “The powerful combination of peer pressure and rationalised rules in the concertive system creates a new iron cage whose bars are almost invisible to the workers it

incarcerates” (Barker, 1993, p. 435). This description of self-management is thoroughly consistent with Foucault’s elaboration of the increasing invisibility of governmentality in modern society (Foucault, 1977; Gore, 1998). Through the panoptic operation of surveillance, which is invisible but constantly possible, self-managing teams and schools have an impact upon specific practices enacted by teachers, but these mechanisms are not obvious either to the teachers themselves, nor, I would argue, to researchers using other methodological approaches.

The major effect of the panopticon in self-managing schools is to

induce in the inmate [teacher] a state of conscious and permanent visibility that assures the automatic functioning of power. So to arrange things that the surveillance is permanent in its effects, even if it is discontinuous in its action; that the perfection of power should tend to render its actual exercise unnecessary; that this architectural apparatus should be a machine for creating and sustaining a power relation independent of the person who exercises it; in short, that the inmates should be caught up in a power situation of which they are themselves the bearers.... The principle [was] that power should be visible and unverifiable. Visible: the inmate will constantly have before his eyes the tall outline of the central tower from which he is spied upon. Unverifiable: the inmate must never know whether he is being looked at at any one moment; but he must be sure that he may always be so. (Foucault, 1977, p. 201)

In the context of schooling in New Zealand at the time this study was undertaken, teachers knew they were being watched. There were countless reports in the media about failing schools – schools that had not passed muster with the ERO. Through appraisal schemes, assessment documentation, effectiveness reviews and reports to the board of trustees their work was constantly visible to others in the system. But power was also unverifiable and unpredictable. Although the ERO visited schools three-yearly, a school would not know very far in advance when exactly the review would happen. As Rose stated, she never knew if she was doing enough in terms of documentation: no one told her exactly what she should do nor how she should do it. Schools were also in this position with an ERO review. They were never quite sure what was required in terms of

assessment and so kept a weather eye on other schools to see what they produced that met with the approval of their ERO review.

In self-managing schools, and the teams within them, teachers worked together to develop norms and rational rules that created micro-level disciplines that rationalised their assessment behaviours so as to make them (seem) purposeful, functional and controlled (Foucault, 1977). As McNeil (1986) explained, in serving the social-control function, the teachers themselves both transmitting and being acted upon by power, become part of the process by which the young are disciplined, and they themselves are controlled by the same forces.

Accountability to colleagues in schools, therefore, is a powerful form of control. It is not that teachers are directed to use checklists and quasi-mastery assessment in which the degree of mastery is rated on a three- in some cases, or four-point scale. As indicated at the beginning of this chapter, the ERO, which many teachers stated they were carrying this type of assessment out for, criticised this approach as inadequate and unsuitable for the evaluation of student progress (Education Review Office, 1999, pp. 22-3); and the Ministry of Education always intended that the primary purpose of assessment should be to improve learning and the quality of learning programmes (Ministry of Education, 1993a). Rather, power is exercised in self-managing schools as disciplinary power (Foucault, 1977), which derives from the use of simple instruments such as “hierarchical observation, normalising judgement and their combination in a procedure that is specific to it, the examination” (p. 170). Conformity, thus, is as much constituted by the teachers as it is constitutive of their practice. In this sense power is not the result of overt force that visibly bends the will of those subject to its operation; instead, conformity results from the constant working of invisible constraints that bring us all toward the same “normal” range of practices and beliefs (Covalesskie, 1993).

As other authors have argued, national and regional “high stakes” that demand assessment for measuring school performance create pressures, or in Woods’ (1997) terms, constraints, that drive teachers to more summative practices (Black, 1998; Broadfoot, 2000; Nisbet, 1993). But, as other literature

demonstrates, it is widely known and understood that it is formative assessment that has the most potential to raise standards and improve learning (for example, Black & Wiliam, 1998b; Crooks, 1988; Gipps, 1994a). The evidence from this research demonstrated how teachers, who valued formative assessment highly, were disciplined through self-management and collaboration to use codes, categories and close observation of their students in terms of the hierarchically organised achievement objectives to characterise much of their assessment. The assessment frenzy that had prompted my research turned out to be teachers' own determination to make the managerial demands upon them to produce evidence of learning manageable and a routine part of their practice.

Some of the teachers in this study described how these instruments operated in the school context as appraisals that checked up on how they were collecting assessment information, while others talked of supervising themselves. These appraisals were, in Foucault's terms, an "examination". For those teachers who used a unit assessment approach, the power of team and self-surveillance seemed most apparent. For example, Rose explained that during these appraisals by senior teachers in her team, she needed to show how she was recording progress using the three-point scale she had described to me, and that she had collected the running records, writing samples and other agreed-upon evidence of learning. Working in self-managing schools meant that teachers were constantly under the surveillance of their peers and themselves. Rose's frustration about not being able to record all she should can be seen as part of this examination "that places individuals in a field of surveillance [and that]also situates them in a network of writing; it engages them in a whole mass of documents that capture and fix them" (Foucault, 1977, p. 189).

The combination of peer pressure and rational rules (Barker, 1993) with the need to meet the requirements of the NEG's and the NAG's in manageable ways, along with, for many teachers, a lack of an in-depth knowledge of the impact of assessment practices on learning (Black & Wiliam, 1998b; Crooks, 1988), has led to the establishment and perpetuation of such quasi-mastery systems of assessment in teachers' everyday practice. The evidence from the research presented here is

that many of the schools where teachers were interviewed have moved in the direction of quasi-mastery systems of assessment, often carried out as some form of unit assessment, as a school or syndicate policy. The ERO (1999), Thrupp et al. (2000) and Dixon (1999) all support this evidence. The potential for systems such as this to dominate teachers' practice is a worrying feature of this research. As Rose's portrait revealed, such assessment systems can divert the teacher from more formative practice and, in more extreme situations, such as that described by Sam, lead to burnout and to leaving teaching completely. Further, as Broadfoot (2000) argues, "the most important concern of all...concerns issues of the most fundamental kind regarding the type of society to which we aspire and the part that education needs to play in building it" (p. 214). Rather than assessment being simply a means of achieving education efficiently, such an obsession with measurement has tended to make efficiency into *the* goal of education (Broadfoot, 2000).

This rationalising power of science, through the evolution of the assessment and measurement discourses addressed in this thesis, has the potential to all but exclude the more fundamental questions about the goals of education itself. Broadfoot cites Charles Darwin as an early victim of this rationalising power of science. According to his autobiography, Darwin stated that he had lost his formerly intense pleasure and desire for art and music through the way he had disciplined his mind to become "a sort of machine for grinding general laws out of large collections of fact" (Broadfoot, 2000, p. 214). Broadfoot argues that the magic and mystery of emotion which can bring the whole process of education alive runs the grave risk of being "subsumed to the necessity of covering the syllabus and doing what is required for the exam" (p. 214).

### Making the "Examination" Formative at the School Level

In the ten years since the administrative and curriculum changes to education it has been teachers, first and foremost, who have had to deal with the competing assessment discourses highlighted in this thesis. The frustration expressed by many teachers was related to their need to keep up with recording

assessment in line with school policy while, at the same time, operating their teaching according to the learning progressions they had in their minds. From these results it seems that teachers would appreciate

- help to make connections between their own tacit craft knowledge and the curriculum objectives
- the re-establishment of formative assessment as being highly valued in learning and teaching
- recognition of the importance of their professional judgement and intuition
- the development of non-intrusive ways of recording and reporting learning outcomes.

Ministry of Education Assessment for Better Learning contracts and teacher education courses are addressing these issues. But as well as professional development and teacher education, this study reveals that syndicate and school-wide policies heavily influence teachers' practices. School policies that build in assessment strategies known to provide the feedback crucial to student learning would seem to offer a more fruitful approach to improving learning and the quality of learning programmes. More specifically, self-managing schools could encourage teacher practices that lead to enhancing learning in several ways. Rather than monitoring achievement against the achievement objectives in every curriculum area, schools could assist teachers to focus on priorities, such as literacy and numeracy, by developing policies that require information about these areas to be collected, used for improving teaching and reported. The revised National Education Guidelines (Ministry of Education, 1999b) may assist in this process.

School guidelines about what assessment data teachers need to collect should be clear and not dominate the classroom programme, nor require teachers to check progress against every achievement objective. Periodic monitoring of important indicators of learning should be enough to demonstrate progress. Schools should encourage the use of assessment tasks that provide information that can be used to diagnose strengths and needs, and provide achievement data.



For example, the diagnostic survey for six-year-olds (Clay, 1979), items from the Assessment Resource Banks (Croft, 1999), the National Education Monitoring Project (Crooks, 1999; Flockton, 1999) and the forthcoming national exemplars and literacy and numeracy tools (Ministry of Education, 1999a) might provide information that can be used for both purposes.

In order to make policies that support teachers in enhancing learning rather than assessing mainly for accountability reasons, it is also critical that school leaders have enough knowledge about assessment and its impacts on teaching and learning. Often, in order to meet managerial demands, school leaders are instrumental in developing policies about school-wide assessment that increase the amount of testing, checklisting and recording. Even though some school leaders read research reports and realise that formative assessment is vital for enhancing learning, their lack of knowledge and understanding about what formative assessment is can lead to management practices that mitigate against its implementation. This can be exacerbated by the software systems for analysing and reporting assessment information. Many of these require checklisted information to be entered in order to produce graphs, charts and tables. To make matters worse, teachers are often expected to enter the assessment data electronically themselves, often on slow hardware systems. This takes more time away from planning and teaching. While computerised systems appear to save time by producing graphs and charts for reporting assessment information, or well-presented reports for parents, they usually add time to teachers' out-of-classroom responsibilities.

The aim is to have policies that encourage teachers to use the assessment information they collect to enhance teaching and learning. Often this information can also be used for other purposes. For example, diagnostic information from running records should primarily influence what the teacher does in the classroom programme to improve reading behaviour. But, with careful planning, it can also be used from time to time to gain syndicate or school-wide assessment data about reading. Portfolio assessment approaches can also provide assessment information that is rich enough to be used for a range of purposes. Supported by benchmark

portfolios that document school-wide expectations for progress in learning in particular areas, teachers can come to hold shared ideas about learning progressions and expectations at different levels of the curriculum. These internalised progressions and expectations allow teachers to work towards raising standards through formative assessment practice on a daily basis, noticing significant learning, acting on it appropriately and recording it if necessary.

One of the strongest influences on teachers' assessment practice, apart from the way in which teachers are required to collect and record assessment information, is the way teachers themselves are assessed. Appraisal and performance management practices should reinforce assessment practices that enhance learning, as opposed to those that engender testing and checklisting.

### **The Impact of Research on Practice**

In line with the theoretical views outlined in the methodology chapter, this research impacted on practice as it occurred. For example, after Rose had read the portrait written about her in Chapter Eight, she commented that she had made significant changes to her practice in line with using formative assessment more consciously and productively; and, as outlined in Chapter Eight, Luke found more systematic ways to record significant learning and act on it during teaching. He asked me for teachers that he could visit to observe how they managed the recording aspects of teaching while prioritising formative assessment.

Both Arawhata and Waimaka Schools changed their policies and assessment systems. At Arawhata School, the roll books described in this thesis became a consistent aspect of teachers' practice, and, in Anne's syndicate at least, have become aids to both formative and summative assessment. The principal and assistant principal retired towards the end of the fieldwork and a new principal and assistant principal were appointed. Communication with the school since the fieldwork ended has suggested that as a school they are considering the implementation of electronic record systems. Anne has continued her contact with the university and has asked for support to sustain her action research into

integrated approaches to teaching and learning. At the time this thesis was completed, she and Rose, who had become a syndicate leader in the Year 5 and 6 area of the school, were considering extending this initiative to the senior school. Rose and Chris, along with two other teachers from Arawhata, have undertaken a university course in assessment.

At Waimaka School, Katy continued as the principal and led the constant revision of the assessment systems there. As well as the student portfolios, Waimaka designed a useful and straightforward record system that met the needs of the teachers, management systems and received an outstanding ERO report in 2000. In order to monitor performance in the school, the teachers at Waimaka used tasks from the National Education Monitoring Project and compared the results of their students with the national picture, especially in mathematics and social studies. Although there has been a complete turnover of staff except for Katy, progress has continued. Katy completed a university course in assessment in 1999 and reported that her involvement in this research process as well as the academic course had increased her confidence and knowledge, and that both had impacted upon her practice and ability to lead assessment to improve learning in the school.

In my own practice through directing the Assessment for Better Learning contract, we gave much attention to enhancing lead<sup>28</sup> teachers' views about how assessment can improve learning through formative assessment, and assisted them to understand better what formative assessment is and how it can be promoted. As a result of the Assessment for Better Learning contract, one school for which we have evidence completely remodelled its systems of assessment, appraisal and performance management in line with strategies that promoted teachers' knowledge and understanding of learning progressions in various areas and their formative assessment practice. This was achieved through the formative use of benchmark portfolios using exemplars of children's work and the production by the teachers of portfolios of evidence about their use of formative assessment to

---

<sup>28</sup> "Lead" teachers attended the professional development sessions and then worked with the rest of the teachers in their respective schools to improve assessment systems and practices.

improve learning and raise achievement in their classes and syndicates. This school, having taken on board the messages about assessment and collaborative professional development, used disciplinary power to move its teachers towards the use of assessment for improving learning while maintaining the ability to check up on outcomes from time to time.

However, it is interesting that the principal of this school stated that this was a “dangerous activity”. By this she meant that because the school had previously received an excellent ERO report for a system of assessment that, although well documented, disciplined teachers to check off the achievement of children against every achievement objective, moving away from this approach to a radically different one opened her and her staff up to the possibility that they might receive a report that contained recommendations for improvement. In the context of self-managing schools in New Zealand, receiving recommendations from the ERO, when there had been none following a previous review, meant that the school had “slipped”. While the principal and teachers in this school were confident to argue for the changes they had made, and to demonstrate the improvements in achievement, a report containing recommendations holds the threat of publication, reaction by the community and the possible loss of students, which, in extreme cases, can mean staff reductions. In the event this school received another particularly positive report from the next ERO review and, thus, were reassured that the ERO valued their new approach. Less confident schools, however, may not be so keen to take the risk, and therefore stick to what they know they have been rewarded for in the past – that is, summative documentation, checklists, collections of children’s work and graphs of achievement based on the aggregated material from the checklists.

### National Policies of Assessment and their Impact on Assessment, Teaching and Learning

The potential for the research reported here to impact on practice is wider than the impact that has already occurred, however. These results have implications for the way that both national and local policies about assessment are

administered. While teacher education and education officials have a responsibility to promote the best practices in assessment, the research reported here indicates that national policies, and school expectations, policies and procedures, are very powerful in determining teachers' assessment practices. Teachers need to be confident about assessment, its terminology and its use so that they can make informed decisions about their classroom practice, prioritise formative strategies over accountability ones when appropriate, and see how summative and standardised assessment might work for, rather than against, learning. School leaders, who usually make the critical decisions about school policies and procedures, could assist teachers in raising standards through assessment by constructing school assessment policies that do not pressure teachers into gathering summative data at the expense of using formative assessment practices.

The findings above, and other evidence (for example, Gilmore, 1998), indicate that while teachers would appreciate help to improve their existing assessment practices, it seems clear that increasing accountability measures further will do little to improve standards. Even teachers with the best knowledge and understanding possible cannot resist the performance imperative when the stakes are raised for purposes of comparison. If we are serious about raising standards in education, the evidence points strongly towards the development of formative assessment in classrooms, because "we know of no other way of raising standards for which such a strong *prima facie* case can be made on the basis of evidence of such large learning gains" (Black & Wiliam, 1998b, p. 19). The challenge for New Zealand schools is to construct policies and procedures that promote teachers' use of assessment to raise standards and improve student achievement over the need to continually provide accountability data.

As this thesis is being completed, new national developments in assessment policy are taking place. Late in 1999 the Ministry of Education published a revised set of National Education Guidelines, significantly different from the NEGs and NAGs in effect during the process of the research reported in this thesis (Ministry of Education, 1999b). There is no doubt that these will

refocus the efforts of schools and their assessment practices. It will take some time to see exactly what the effects of the revised National Education Guidelines will be. The rewording to include such statements as “giving priority to student achievement in literacy and numeracy, especially in Years 1-4” (NAG 1 (i)) and “giving priority first to student achievement in literacy and numeracy, especially in Years 1-4; and then to breadth and depth of learning...” (NAG 1 (ii)), may move schools towards requiring more formative practice. The new National Education Guidelines also tighten up on bi-cultural issues. But, in contrast, the fact that many teachers do not understand how assessment can prioritise learning and that, even if they do, their understanding of formative assessment is often as continuous summative, similar to that described as a unit assessment approach in this thesis, these modifications to the National Education Guidelines may have little effect.

In addition, the proposals to provide tools to measure success in literacy and numeracy at nine (Gerritsen, 2000) and to develop exemplars of students’ work across the curriculum linked to levels in the national curriculum, bring with them the danger that instead of focusing schools and teachers on improving their formative assessment practices and using summative assessment sparingly to provide snapshots of progress when appropriate, schools may use these new devices to measure outcomes continually. For example, having exemplars of student work linked to levels of the curriculum developed to reveal national standards for the achievement objectives at curriculum levels 1 to 5 may well encourage schools to use these as benchmarks to be achieved at the end of each unit, to aggregate the results school-wide and to discipline teachers to use them as the framework for their assessment and reporting. Should the ERO reinforce this with recommendations and compliance requirements in their reports, and the community look for progress stated against levels as used in the exemplars, the potential exists for a greater move than at present towards summative assessment at the expense of more formative strategies. While it must be emphasised the above scenario is not the intention, the indications from this research about how teachers behave in the context of self-managing schools under the supervision of

the ERO and the surveillance of their community suggest that the reality has the potential to be somewhat different.

### Impacts on My Own Practice

As a result of undertaking this research, I have come to a fresh understanding of assessment and the contexts of schooling and teaching in New Zealand. Formerly I had a somewhat narrower view of both assessment and its place in the school system. I am now vitally aware of the competing discourses of assessment, learning and education within society and over time, and the impacts these have on individuals and schools, as well as upon teacher education, the Ministry of Education and agencies such as the Education Review Office. Using Foucault's tools, such as discourse, power/knowledge, panoptism, hierarchical judgement, normalisation and the examination, assisted me to see beyond technical understandings of assessment terminology and understand the effects that the restructuring of education had upon assessment and teachers' assessment practices in the 1990s.

Through the use of these tools I was able to understand why formative assessment is so powerful in facilitating learning; how the teachers' constant gaze, focused upon significant and valued learning, disciplines learners and, in turn, causes them to value and focus upon certain learning. I was also able to see how certain knowledge in our society is valued and reinforced through this process while other knowledge is marginalised and excluded. For example, in specifying particular achievement objectives as important, say in literacy and numeracy in English for children in mainstream schools, others that may be valued by minority sectors of society, such as rote learning and learning by heart are marginalised.

The notion of panoptism enabled me to understand better why schools perceived the ERO to be such a threat and how making schools self-managing and open to the gaze of their community through the governance of the boards of trustees was so powerful in promoting particular assessment practices above others, usually summative above formative. In fact, in the climate of suspicion about teachers generated shortly before and during the introduction of the

Tomorrow's Schools restructuring, there was a downplaying of what has often been referred to as teachers' professional judgement. From the results of this research it would seem that teachers' professional judgement underpins and informs (and may even be a synonym for) formative assessment. Foucault believed that

to combat effectively the carceral society's infinitely minute web of panoptic techniques one needs to acknowledge that contemporary (post) industrial society is characterised by radically diffuse and localised power relations and that the source and mechanisms of power are not to be found in any one site. (Blacker, 1998, p. 356)

Therefore, concentrating on power's "microphysics" from the bottom up – in other words, from its effects upon teachers in this case – allowed me to look with a different focus at how managerial discourses were influencing teachers' practices and why these discourses were so powerful. These understandings caused me to reconsider my role as a researcher and teacher educator. Instead of espousing truth about assessment, or teaching and learning, for example, I became intrigued to know more about "what what we do does". To use Foucault's explanation from an interview in 1982, "people know what they do; they frequently know why they do what they do; but what they don't know is what what they do does" (Foucault, 1982, p. 187). Or as Blacker (1998) explains, "by restricting one's scope of activity, one often widens and deepens the (potential) consequences of that activity" (p. 359).

In particular, this perspective helped me to appreciate how the structure of the education system, which uses the device of self-managing schools labelled with deciles by their socio-economic status, sets educational standards as those results achieved mainly by students in decile 10 (the highest socio economic group). To "close the gap" (Gerritsen, 2000; Velde, 2000) between the achievement of children in decile one schools and those in decile 10 schools seems, by definition, almost impossible. Success in our society is measured in socio-economic terms, thus decile 10 schools contain children from successful families. This was brought home to me by a statement from a teacher in a decile 1



school who told me how pleased they are when, very occasionally, a decile 10 school receives a poor ERO report.

As well as socio-economic differences perpetuated through dominant discourses, gender, cultural and ethnic disadvantages are perpetuated through schooling. Bishop and Glynn (1999) argue that there has been a cumulative development of cultural superiority in New Zealand and that New Right market philosophies have continued to exacerbate notions of superiority through the administrative and curricula reforms within education. As I have noted throughout this thesis, there is a “silence” of culture in the assessment discourses as well as in the research literature and methodology arguments. The use of the Foucauldian “toolkit” in this thesis has, for me, had the advantage of disrupting accepted rationalities about assessment so that I can see how having one set of achievement objectives for all learners in New Zealand on which they are assessed and by which they become “known” as educated is, in and of itself, disadvantaging. In the present standards-based discourse of assessment there appears to be little room for diversity. The “gaps” in the system are created by distance from the norm. Because the values and objectives of the school-based discourses are those of the dominant pakeha culture, which, by its very nature, treats Māori and minority cultures as “other” and their social practices as “deviant” and “non-standard”, Māori must either become complicit with values that denigrate and damage their culturally-based discourses and identity or remain as “other” and as “non-achievers”.

The three teachers’ stories, as well as conveying the power of conflicting discourses, showed that all three teachers were doing more assessment-in-action than would have been visible from the interviews I conducted. This finding demonstrates the vital importance of using multiple data sources; in this case, observation, amplified and extended data from repeated interviews, and the assessment records of teachers. These sources enabled a better understanding and appreciation of how the “examination” works in a standards-based system, as well as an appreciation of how teachers, in effect, produce the “gap”. This has led me to consider the need for completely different administrative and curriculum

structures that include diverse discourses and standards. In a market model, the checks and balances of the system will always work against those people classified as “other”. “Closing the gaps”, therefore, may require redefinition of the curriculum and the way it is assessed in ways that include and value more diverse discourses.

### **Limitations of the Study and Areas for Further Research**

Being a study of singularity, a participative case study such as that used in this investigation is limited in the extent to which generalisations can be made on the basis of the results. While sufficient data were collected to be able to explore and describe teachers’ different approaches to assessment and the effect that multiple competing discourses had upon their assessment practice, it is not possible to make sweeping generalisations. However, as Bassey suggested, qualitative case studies can produce what he terms “fuzzy generalisations” (1999, p. 51). Fuzzy generalisations in his terms are general statements with built-in uncertainty. As he explains

With the scientific generalization there are no exceptions – and indeed in science if any are found then the statement is abandoned or revised to accommodate the new evidence. But in the use of the adjective “fuzzy” the likelihood of there being exceptions is clearly recognized and this seems an appropriate concept for research in areas like education where human complexity is paramount. (Bassey, 1999, p. 52)

The reasons for using fuzzy generalisations include that they are firm reminders that the determinants of learning are complex and that such tentative statements tend to invite teachers and others to enter the debate.

Rather than see this as a limitation in a praxis-oriented approach such as the one employed in this investigation, the quest is to open up spaces for new practices. The limitations here may be, rather, the narrowness of interpretation of assessment practice in terms of three approaches – head noting, unit assessment and an integrated systematic type. In fact, categorising assessment at all is somewhat problematic. As the portraits in Chapter 8 revealed, each of these three

teachers used all of the so-called approaches at one time or another. It should be noted that the categories used related to the predominant use of one type of assessment and recording strategy as described by the teachers themselves rather than to the teachers. During the progress of the fieldwork I deliberated at length about whether or not to use these or any categories as a tool for analysis. The main reasons for proceeding with their use were that they provided a way to differentiate between the descriptions of practice offered by the teachers, enabled some way of selecting a small number of teachers for further study and established a framework that served as a tool for talking with teachers in more depth about assessment.

However, the worth of categorising assessment behaviour became much less clear after the classroom observations. It became apparent that what the literature calls formative assessment was often thought of in other terms by these teachers. All three made substantial use of “head noting” in their daily practice, but some did not mention this when asked about their assessment. It may well have been possible to use categories based on other aspects of assessment practice, such as types of feedback given or on what happened in different curriculum areas. Alternatively, a narrative approach that did not use categories at all, such as that used for the portraits in Chapter 8, could have been employed. While the use of categories can be seen as a limitation of this study, they were extremely useful as a tool to get teachers talking about their practice in more detail than was possible during the first two rounds of interviewing.

Within a singular curriculum that provides one view of knowledge – that of a hierarchy of achievement objectives through which all students must be known – it is tempting to provide specific answers to the question of how learning may be improved. I believe that, although the use of Foucault’s interpretive analytics had the advantage of disrupting accepted rationalities and taken-for-granted claims and plausibility structures about assessment, and raised questions about the contemporary discursive practices, the conclusions reached were still narrow in their scope. They were narrow first in their mono-cultural views of assessment. At no time did I attempt to look for approaches to assessment that may be more

appropriate to raising the achievement of Māori or minority migrant or refugee groups. It could even be argued that this thesis excluded or marginalised particular assessment strategies, such as the assessment of groups in contrast with individuals, or the use of peer and self-assessment in contrast with teacher assessment, which might hold promise as strategies by which standards could be raised.

Rather than these issues being seen as limitations of this study, they should be the focus of future investigations. The positioning of Māori and other “disadvantaged” groups, which at times is distinctly oppositional, is an area for further investigation, since cultured subjectivity presents a conception of power which relies on knowledge to fix what is normal. Particularly in literacy and numeracy, Māori and other cultural groups are placed differentially, oppositionally and unequally to mainstream pakeha students. In a society where knowledge is commodified by the dominant discourses and its achievement supervised by the community through both local board of trustees and national (ERO) mechanisms, there may be a need to broaden and deepen teachers’ knowledge of alternative and diverse discourses in order that those other than the norm may find a place. Further research of the ways in which assessment excludes and marginalises the achievement and progress of “disadvantaged” and culturally different groups may assist.

During the conduct of this investigation it became clear that the use of electronic technology for assessment, recording and reporting purposes was increasing dramatically. The Ministry of Education launched a technology initiative to assist schools in this respect. This investigation, apart from acknowledging the uses of this type of technology, did not focus directly upon it. The use of technology will influence teachers’ practices heavily, even within the near future, and has the potential to drive assessment in either formative or summative directions. Investigations as to how this type of technology can be used to support learning and achievement rather than promote more summative checklisting approaches to assessment would be a positive development.

Although a variety of positive impacts on practice from this investigation were listed earlier in this chapter, the major outcomes of the research itself have been the production of this thesis and various conference and journal papers. From my perspective as a teacher educator first and foremost, I see this as a limitation. While an action research approach to this research may have been more productive in terms of improving the assessment practices of teachers in the two case study schools, the more theoretical approach taken produced new ways of understanding the effects of the administrative and curriculum reforms of the early 1990s on teachers' practice. Rather than applying theory to practice, the methodology in this study has developed theory from practice. In contrast to action research, which is concerned with contributing to the development of the case or cases under study by feedback of information which can guide revision and refinement of the action (Stenhouse, 1985, cited in Bassey, 1999), this investigation has attempted a new type of academic classroom research which investigates teaching at its source with the complicity of teachers but is not action research.

A further possible limitation of this study may have been my multiple roles during its conduct. Although, primarily, my role was that of researcher, I was also a consultant at the two case study schools in the first year or so of the investigation, I carried out action research for the Ministry of Education in the schools in the national sample, and I was well known in the Waikato region as a teacher educator with a particular interest in the assessment area. Participants from the case study schools, particularly Rose, Luke and Anne, were asked how these multiple roles might have impacted on their responses or on their teaching. Although they did not report any obvious impacts it is still possible, and indeed likely, that their actions and responses were influenced by these multiple positionings.

The results of this type of research are helpful in making teachers' tacit knowledge explicit and assisting teachers to gain voice for their tacit knowledge – voice that can then have more of a say in how, for example, assessment might improve learning. Using Foucault's tools of disciplinary power, discourse and the

notion of assessment as “the examination” has demonstrated how things might be different. There is scope for further research into how self-management mechanisms increase self- and peer surveillance for the implementation of national and school policies and practices; for example, understanding how disciplinary power works in self-managing schools to influence teachers’ planning and curriculum choices. It is also now timely to work in partnership with schools to assist them to use the results of this thesis to extend teachers’ and school leaders’ knowledge about assessment in general, and to support them in using this knowledge to inform their management, teaching and assessment practices to improve learning.

## **Conclusion**

If we are serious about raising standards in education, we should accept that the evidence points strongly towards the need to develop formative assessment in classrooms, because as already stated “we know of no other way of raising standards for which such a strong *prima facie* case can be made on the basis of evidence of such large learning gains” (Black & Wiliam, 1998b, p. 19). The challenge for New Zealand schools is to construct policies and procedures that promote teachers’ use of assessment to raise standards, improve and enthuse student learning over the need to continually provide achievement data for accountability purposes. This will take a shift in the discourse of education and assessment throughout the system.

Formative assessment practices built upon teachers’ knowledge of progressions in learning, however, will be promoted only when they are expected and rewarded by the very mechanisms which currently promote summative assessment practices. These mechanisms include, but may not be limited to, surveillance of formative practice by the ERO, and school policies and practices, and appraisal and performance management systems that encourage formative approaches through supporting teachers to develop their mental “maps and timelines” in particular areas of learning in contrast to dominant accountability

mechanisms. As educators we have become obsessed with measuring individuals' competence against a nationally prescribed progression of achievement objectives. Although attempts to introduce national externally referenced tests for all students at particular age levels in primary schools have to date been resisted, other new assessment initiatives have the potential to discipline teachers to invent new forms of summative practice. It is my hope that this thesis will assist people at all levels of the education system to understand more about how policies, both at the national and local (school) level come together to discipline teachers into particular classroom assessment practices that have powerful impacts, which may be positive or negative, on learners. By knowing more about these effects, and how and why they occur, we as teachers, teacher educators, school leaders, academics and education officials may be in a better position to understand what "what we do" does and to see that within our own sphere of interest we can make a difference.

## REFERENCES

- Adams, J. C. (1885). Examinations and tests for schools. The New Zealand Schoolmaster, IV(8), 124.
- Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 377-392). Thousand Oaks, CA: Sage Publications.
- Afflerbach, P., & Johnston, P. (1993). Writing language arts report cards: Eleven teachers' conflicts of knowing and communicating. The Elementary School Journal, 94(1), 73-86.
- Aikin, S. (1994). Primary problems and the New Zealand Curriculum Framework. New Zealand Annual Review of Education, 4, 57-76.
- Anders, P., & Richardson, V. (1994). Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions and consequences of accountability. Teachers' College Record, 94(2), 382-396.
- Apple, M. (1992). Ideology and the national curriculum. Paper presented at the John Dewey Lecture, John Dewey Society and the American Association for Research in Education, San Francisco.
- Apple, M. (1993). Official knowledge. New York: Routledge.
- Apple, M. (1996). Cultural politics and education. Buckingham: Open University Press.
- Apple, M. (1998). Foreword. In S. Middleton (Ed.), Disciplining sexuality: Foucault, life histories, and education (pp. vii-xi). New York: Teachers' College Press.
- Argyris, C. (1993). Knowledge for action: A guide for overcoming barriers to organizational change. San Francisco: Jossey-Bass.
- Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. Applied Measurement in Education, 4(4), 275-288.
- Austin, M., Edwards, W., & Parata-Blane, A. (1997). Review of the educational external evaluation system: Background and issues paper. Wellington, New Zealand: Author.



Bachor, D. G., & Anderson, J. O. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. Assessment in Education: Principles, Policy and Practice, 1(1), 63-93.

Ball, D. G. (1948). What does integration mean? Education, 1(3), 115.

Ball, S. J. (1985). Participant observation with pupils. In R. G. Burgess (Ed.), Strategies of educational research (pp. 23-25). London: The Falmer Press.

Ball, S. J. (Ed.). (1990). Foucault and education: Disciplines and knowledge. London: Routledge.

Barker, J. R. (1993). Tightening the iron cage: Concertive control in self-managing teams. Administrative Science Quarterly, 38, 408-437.

Bassey, M. (1999). Case study research in educational settings. Buckingham: Open University Press.

Battersby, D. (1981). The first year of teaching: A grounded theory study. Unpublished doctoral thesis, University of Waikato, Hamilton, New Zealand.

Beane, J. (1992). Creating an integrative curriculum: Making the connections. Curriculum Theory and Practice, NASSP Bulletin(November), 46-54.

Beattie, M. (1997, June). Redressing the balance in educational research: The possibilities inherent in narrative inquiry for connecting curriculum development, professional development, and educational reform. Paper presented at the Biennial Conference of the Australian Curriculum Studies Association.

Beeby, C. E. (1986). Introduction. In W. Renwick (Ed.), Moving targets (pp. xi-xlv). Wellington, New Zealand: New Zealand Council for Educational Research.

Beeby, C. E. (1992). The biography of an idea: Beeby on education. Wellington, New Zealand: New Zealand Council for Educational Research.

Bell, B., & Cowie, B. (1997). Formative assessment and science education: Summary report of the learning in science project (assessment) (Research Report for the Ministry of Education). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Benton, L. M. (1999). Environmental discourse and practice. Oxford: Blackwell Publisher.

Bernstein, B. (1971). On the classification and framing of educational knowledge. In M. Young (Ed.), Knowledge and control: New directions for the sociology of education (pp. 47-69). Stockley Close, West Drayton, Middlesex, U.K.: Collier MacMillan.

Bishop, R., & Glynn, T. (1999). Culture counts: Changing power relations in education. Palmerston North, New Zealand: The Dunmore Press.

Black, P. (1998). Testing: Friend or foe? Theory and practice of assessment and testing. London: Falmer Press.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. London: School of Education, Kings' College.

Blacker, D. (1998). Intellectuals at work and in power: Toward a Foucauldian research ethic. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (pp. 348-368). New York: Teacher's College Press.

Blackmore, J. (1988). Assessment and accountability. Geelong, Victoria: Deakin University Press.

Bogdan, R. C., & Biklen, S. K. (1982). Qualitative research for education: An introduction to theory and methods. Boston, MA: Allyn and Bacon.

Borko, H., & Elliott, R. (1998, June). Tensions between competing pedagogical and accountability commitments for exemplary teachers of mathematics in Kentucky. Paper presented at the National Conference on Large-Scale Assessment of the Council of Chief School Officers, Colorado Springs, CO.

Boston, J. (1991). The theoretical underpinnings of public sector restructuring. In J. Boston, J. Martin, J. Pallot, & P. Walsh (Eds.), Reshaping the state: New Zealand's bureaucratic revolution (pp. 1-26). Auckland, New Zealand: Oxford University Press.

Bourke, R., & Willis, D. (1998). Teachers' use of assessment: Changing practices, changing minds . Unpublished paper.

Broadfoot, P. (1995). Performance assessment in perspective: International trends and current English experience. In H. Torrance (Ed.), Evaluating authentic assessment (pp. 9-43). Buckingham, U.K.: Open University Press.

Broadfoot, P. (1996). Education, assessment and society. Buckingham, U.K.: Open University Press.

Broadfoot, P. (2000). Assessment and intuition. In T. Atkinson & G. Claxton (Eds.), The intuitive practitioner: On the value of not always knowing what one is doing. (pp. 199-219). Buckingham, U.K.: Open University Press.

Broadfoot, P., & Pollard, A. (1998, 13-17 April). Categories, standards and instrumentalism: Theorising the changing discourse of assessment policy in English primary education. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Brodhagen, B. L. (1998, 13-17 April). Teacher decision making within curriculum integration. Paper presented at the Annual Meeting of the American Education Research Association, San Diego.

Brown, M. (no date). Formative assessment for learning: General issues illustrated by examples from England . (Unpublished paper)

Brown, S. (1996, 28-30 June). Keynote address. Paper presented at the Symposium on Validity in Educational Assessment, University of Otago, New Zealand.

Bruce Ferguson, P. (1999). Developing a research culture in a polytechnic: An action research case study. Unpublished doctoral thesis, University of Waikato, Hamilton, New Zealand.

Bruner, J. (1960). The process of education. Cambridge, MA: Harvard University Press.

Burt, C. (1948). The subnormal child. Education, 1(1), 39-42.

Butchers, A. G. (1932). The education system: A concise history of the New Zealand education system. Auckland, New Zealand: The National Printing Company.

Calder, I. (1995, 5-8 July). The impact of changes to assessment: Accountability gone mad? Paper presented at the Australian Teacher Education Association, Australian Catholic University, Sydney, Australia.

Calder, I. (1996). Assessment for better learning (Final report to the Ministry of Education on a contract for the delivery of a professional development programme). Hamilton: University of Waikato.

Carlos, D. (1885). Letter to the Editor. The New Zealand Schoolmaster, IV(7), 124.

Carr, M., & Cowie, B. (1997, 4-7 December). Assessment: Why record? Paper presented at the Annual Conference of the New Zealand Association for Research in Education, University of Auckland, Auckland, New Zealand.

Carr, W., & Kemmis, S. (1986). Becoming critical: Education, knowledge and action research. London: The Falmer Press.

Centre for Graduate Studies School of Education. (2000). A guide to research for graduate students. Hamilton, New Zealand: University of Waikato.

Charmaz, K. (1995). Grounded theory. In J. A. Smith, R. Harre, & R. van Langenhove (Eds.), Rethinking methods in psychology (pp. 27-49). London: Sage Publications.

Clandinin, D. J. (1998, 14-17 October). Stories to live by on the professional knowledge landscape. Paper presented at the New Zealand Council for Teacher Education Conference, University of Waikato, Hamilton, New Zealand.

Clandinin, D. J., & Connelly, F. M. (1995). Teachers' professional knowledge landscapes (Vol. 15). New York: Teachers' College Press.

Clandinin, D. J., & Connelly, F. M. (1996). Teachers' professional knowledge landscapes: Teacher stories-stories of teachers-school stories-stories of schools. Educational Researcher, 25(3), 24-30.

Clark, J. (1998, 14-17 October). ERO's "The capable teacher": An inadequate model for teacher education. Paper presented at the New Zealand Teacher Education Conference, University of Waikato, Hamilton, New Zealand.

Clay, M. M. (1972). Reading: The patterning of complex behaviour. Auckland, New Zealand: Heinemann Educational Books.

Clay, M. M. (1979). The early detection of reading difficulties: A diagnostic survey with recovery procedures. Auckland, New Zealand: Heinemann.

Clay, M. M. (1993). An observation survey of early literacy achievement. Auckland: Heinemann.

Clay, M. M., Gill, M., Glynn, T., & Salmon, K. (1983). Record of oral language and biks and gutches. Auckland, New Zealand: Heinemann.

Cochran-Smith, M., & Lytle, S. L. (1990). Inside/outside: Teacher research and knowledge. New York: Teachers' College Press.

Codd, J. (1994). Educational reform and the contrary discourses of evaluation. Evaluation and Research in Education, 8, 41-54.

Codd, J. (1995). Managerialism: The problem with today's schools. Delta, 44, 17-25.

Codd, J. (1998, 14-17 October). Professional accountability and the education of teachers. Paper presented at the New Zealand Teacher Education Conference, University of Waikato, Hamilton, New Zealand.

Codd, J. (1999). Educational reform, accountability and the culture of distrust. New Zealand Journal of Educational Studies, 34(1), 45-53.

Codd, J., McAlpine, D., & Poskitt, J. (1995). Assessment policies in New Zealand: Educational reform or political agenda? In R. Peddie & B. Tuck (Eds.), Setting the standards (pp. 32-63). Palmerston North, New Zealand: Dunmore Press.

Cohen, L., & Manion, L. (1989). Research methods in education (3rd ed.). London: Routledge.

Comber, B. (1996, 25-29 November). Professional bandwagons and local discursive effects: Reporting the literate student. Paper presented at the Singapore Education Research Association and the Australian Association for Research in Education Joint Conference, Singapore.

Comber, B. (1997). Managerialist discourses: Local effects on teachers' and students' work in literacy lessons. Discourse, 17(3), 389-407.

Comber, B. (no date). Reporting the literate student: Teachers' professional knowledges and the risks for disadvantaged students . (Unpublished paper)

Covaleskie, J. F. (1993). Power goes to school: Teachers, students and discipline. Philosophy of Education. Available: [http://www.ed.uiuc.edu/EPS/PES-Yearbook/93\\_docs/COVALESK.HTM](http://www.ed.uiuc.edu/EPS/PES-Yearbook/93_docs/COVALESK.HTM) [2000, 5 March].

Cowie, B. (1997). Formative assessment and science classrooms. In B. Bell & R. Baker (Eds.), Developing the science curriculum in Aotearoa New Zealand (pp. 115-129). Auckland, New Zealand: Longman.

Crocombe, G., Enright, M., & Porter, M. (1991). Upgrading New Zealand's competitive advantage. Auckland, New Zealand: Oxford University Press.

Croft, C. (1999). School-wide assessment: Using the assessment resource banks. Wellington, New Zealand: New Zealand Council for Educational Research.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58(4), 438-481.

Crooks, T.J. (1999, 19-23 April). Brief overview of the National Education Monitoring Project. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Crossley, M., & Vulliamy, G. (1984). Case-study research methods of comparative education. Comparative Education, 20(2), 193-207.

Crumpler, T. P. (1996). Exploring a culture of assessment with 9th Grade students: Convergences of meaning within dramas of assessment. Unpublished doctoral dissertation, The State University of Ohio, Columbus, Ohio.

Dale, R. (1990, August). The impact of the restructuring of education in New Zealand and Britain. Paper presented at the 2nd Research into Educational Policy Conference, University of Victoria, Wellington, New Zealand.

Darling-Hammond, L. (1997). The right to learn: A blueprint for creating schools that work. San Francisco: Jossey-Bass.

Darling-Hammond, L., & Wise, A. (1985). Beyond standardization: state standards and school improvement. The Elementary School Journal, 85(5), 315-336.

Delandshere, G., & Petrosky, A. R. (1994). Capturing teachers' knowledge: Performance assessment. Educational Researcher, 23(5), 11-18.

Department of Education. (1978). Educational standards in state schools: A report to the Minister of Education. Wellington, New Zealand: Department of Education.

Department of Education. (1986). The curriculum review. Wellington, New Zealand: Department of Education.

Department of Education. (1988). Administering for excellence: Effective administration in education. Wellington: Government Printer.

Department of Education. (1989a). Assessment for better learning. Wellington: Government Printer.

Department of Education. (1989b). Keeping school records: Principles and practice of assessment and evaluation. Wellington, New Zealand: Department of Education.

Department of Education. (1989c). Keeping schools records: Primary progress records. Wellington, New Zealand: Government Printer.

Dixon, H. (1999). The effect of policy on practice: An analysis of teachers' perceptions of school based assessment practice. Unpublished master's thesis, Massey University, Palmerston North, New Zealand.

Dixon-Krauss, L. (1996). Vygotsky in the classroom: Mediated literacy instruction and assessment. White Plains, N.Y., USA: Longman.

Donmoyer, R. (1990). Generalizability and the single case study. In E. W. Eisner & A. Peshkin (Eds.), Qualitative inquiry in education: The continuing debate (pp. 175-200). New York: Teachers' College Press.

Dreyfus, H. L., & Rabinow, P. (1982). Michel Foucault: Beyond structuralism and hermeneutics. Brighton, UK: Harvester Press.

Education Review Office. (1994). Effectiveness Review Report for "Arawhata" Primary School. Hamilton, New Zealand: Education Review Office.

Education Review Office. (1995a). Assessing student achievement (National Education Evaluation Reports 3). Wellington, New Zealand: Education Review Office.

Education Review Office. (1995b). Assurance Audit Report on "Waimaka" Primary School. Hamilton, New Zealand: Education Review Office.

Education Review Office. (1995c). "Barriers to learning" (National Evaluation Reports 9). Wellington, New Zealand: Education Review Office.

Education Review Office. (1999). Student assessment: Practices in primary schools (Education Evaluation Reports 2). Wellington, New Zealand: Education Review Office.

Education Review Office. (no date-a). Accountability in action. Wellington, New Zealand: Education Review Office.

Education Review Office. (no date-b). Evaluating education. Wellington, New Zealand: Education Review Office.

Education Review Office. (no date-c). Evaluation towards effective education: A guide to effectiveness reviews. Wellington, New Zealand: Education Review Office.

Eduvac. (1998, 25 May). Principals against national testing of primary school children. The Education Weekly, pp. 1.

Eisner, E. W., & Peshkin, A. (1990). Introduction. In E. Eisner & A. Peshkin (Eds.), Qualitative inquiry: The continuing debate (pp. 1-13). New York: Teachers' College Press.

Elbaz, F. (1983). Teacher thinking: A study of practical knowledge. London: Croom Helm.

Elley, W. (1993a, 17-20 May). Is standards-based assessment feasible? Some blunt words about some blunt instruments. Paper presented at the National Workshop on Professional Development in Educational Assessment, Stella Maris Conference Centre, Wellington, New Zealand.

Elley, W. (1993b, 19 September). Our curriculums are following a faulted model. Dominion Sunday Times.



Elliott, J. (1991). Action research for educational change. Milton Keynes: Open University Press.

Elliott, J. (1998). The curriculum experiment: Meeting the challenge of social change. Buckingham, UK: Open University Press.

Ewing, J. L. (1960). Origins of the primary school curriculum 1840-78. Wellington, New Zealand: New Zealand Council for Educational Research.

Ewing, J. L. (1970). Development of the New Zealand primary school curriculum 1877-1970. Wellington, New Zealand: New Zealand Council for Educational Research.

Faire, M., Brown, C., & Yates, R. (1994). Recording and reporting: An exploratory study (An exploratory study carried out under contract to the Ministry of Education). Hamilton, New Zealand: University of Waikato.

Faire, M., & Yates, R. (1994). Assessing and evaluating student learning. In C. McGee & D. Fraser (Eds.), The professional practice of teaching (pp. 109-130). Palmerston North, New Zealand: Dunmore Press.

Faulds, G. (1988). Mental testing and social selection: A case study of the use of intelligence and attainment tests to select children for streamed classes at Kowhai Junior High School 1922-1937. In M. Olssen (Ed.), Mental testing in New Zealand: Critical and oppositional perspectives (pp. 90-99). Dunedin: University of Otago Press.

Fenner, L. J. (1995). Student portfolios: A view from inside the classroom. Unpublished doctoral dissertation, The Ohio State University, Columbus, Ohio.

Filer, A. (1995). Teacher assessment: Social process and social product. Assessment in Education: Principles, Policy and Practice, 2(1), 23-38.

Firestone, W., & Winter, J. (1998, 13-17 April). Different policies, common practice: mathematics assessment and teaching in the United States and England and Wales. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Fitzsimons, P. (1996). Governing the self: A Foucauldian critique of managerialism in education. Unpublished doctoral thesis, University of Auckland, Auckland, New Zealand.

Fitzsimons, P., Peters, M., & Roberts, P. (1999). Economics and the educational policy process in New Zealand. New Zealand Journal of Educational Studies, 34( 1), 35-44.

Flockton, L. (1999). School-wide assessment: National education monitoring project. Wellington, New Zealand: New Zealand Council for Educational Research.

Fontana, A., & Frey, J. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 361-376). London: Sage Publications.

Foucault, M. (1977). Discipline and punish: The birth of the prison (Alan Sheridan, Trans.). London: Penguin Books.

Foucault, M. (1978). The history of sexuality (Robert Hurley, Trans.) (Vol. 1). New York: Random House.

Foucault, M. (1980). Prison talk. In C. Gordon (Ed.), Power/knowledge: Selected interviews and other writings by Michel Foucault 1972-1977 (pp. 37-54). New York: Pantheon.

Foucault, M. (1981). Questions of method. Ideology and consciousness, 8, 3-14.

Foucault, M. (1982). Afterword: The subject and power. In H. L. Dreyfus & P. Rabinow (Eds.), Michel Foucault: Beyond structuralism and hermeneutics (1st ed., pp. 208-226). Brighton, Sussex: The Harvester Press.

Fullan, M. (1998). Leadership for the 21st century. Breaking the bonds of dependency. Educational Leadership, 55(7), 6-10.

Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B. R. Gifford & M. C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 77-119). Boston: Kluwer Academic Publishing.

Gardner, H. (1993). The unschooled mind. London: Fontana.

Gee, J. P. (1990). Social linguistics and literacies: Ideology in discourses. London: The Falmer Press.

Gerritsen, J. (2000). New assessment initiatives launched. New Zealand Education Gazette, 79(14), 11-12.

Gifford, B., & O'Connor, M. (1992). Future assessments: Changing views of aptitude, achievement and instruction. Boston: Kluwer Academic.

Gilmore, A. (1998). Assessment for success in primary schools: Report of the submissions to the green paper (Report for the Ministry of Education). Christchurch, New Zealand: University of Canterbury.

Gipps, C. (1994a). Beyond testing: Towards a theory of educational assessment. London: The Falmer Press.

Gipps, C. (1994b). Developments in educational assessment: What makes a good test? Assessment in Education: Principles, Policy and Practice, 1(3), 283-291.

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). Intuition or evidence? Buckingham, U.K.: Open University Press.

Gipps, C., McCallum, B., & Brown, M. (1996). Models of teacher assessment among primary school teachers in England. The Curriculum Journal, 7(2), 167-183.

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory. Chicago: Aldine.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 18, 519-521.

Glaser, R., & Silver, R. (1994). Assessment, testing, and instruction. Review of Research in Education, 20, 393-419.

Glover, P., & Thomas, R. (1999). Coming to grips with continuous assessment. Assessment in Education: Principles, Policy and Practice, 6(1), 117-127.

Gordon, L. (1995). Human capital theory and the death of democracy: Tertiary education policy in New Zealand. In M. Olssen & K. Morris Matthews (Eds.), Education, democracy and reform (pp. 60-73). Auckland, New Zealand: The University of Auckland.

Gore, J. M. (1998). Disciplining bodies: On the continuity of power relations in pedagogy. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (pp. 231-251). New York: Teachers' College Press.

Grace, G. (1990). The New Zealand Treasury and the commodification of education. In S. Middleton, J. Codd, & A. Jones (Eds.), New Zealand educational policy today: Critical perspectives (pp. 27-39). Wellington, New Zealand: Allen and Unwin.

Graves, D. H. (1983). Writing: Teachers and children at work. London: Heinemann Educational Books.

Hakim, C. (1993). Research analysis of administrative records. In M. Hammersley (Ed.), Social research: Philosophy, politics and practice (pp. 131-145). London: Sage Publications.

Hargreaves, A. (1994). Changing teachers, changing times: teachers' work and culture in the postmodern age. London: Cassell.

Hargreaves, A. (1999, 22-24 February). Reinventing professionalism: teacher education and teacher development for a changing world. Paper presented at the International Conference on Teacher Education, Hong Kong Institute of Education.

Hargreaves, A., & Evans, R. (1997). Teachers and educational reform. In A. Hargreaves & R. Evans (Eds.), Beyond educational reform: Bringing teachers back in (pp. 1-18). Buckingham: Open University Press.

Harlen, W. (1998, 3-6 December). Classroom assessment: A dimension of purposes and procedures. Paper presented at the Annual Conference of the New Zealand Association for Research in Education, Dunedin, New Zealand.

Harlen, W., Chisnall, M., & Byrne, M. (1995). Teachers' assessment and national testing in primary schools in Scotland: Roles and relationships. Assessment in Education: Principles, Policy and Practice, 2(2), 129-144.

Harlen, W., Gipps, C., Broadfoot, P., & Nuttall, D. (1992). Assessment and the improvement of education. The Curriculum Journal, 3(3), 215-230.

Harold, B. (1995). Negotiation and contestation during times of change: The introduction of new educational policies to three rural schools. Unpublished D.Phil thesis, University of Waikato, Hamilton, New Zealand.

Harold, B., Hawksworth, L., Mansell, H., & Thrupp, M. (1999). A decade of change and continuity: Emerging themes from the front line of school reform. New Zealand Journal of Educational Studies, 34(1), 234-246.

HarperCollins. (1998). Collins English dictionary: Millenium edition (Millenium ed.). Glasgow: HarperCollins Publishers.

Harvey, D. (1989). The condition of postmodernity. Oxford: Oxford University Press.

Hattie, J., & Jaeger, R. (1998). Assessment and classroom learning: A role for summative assessment. Assessment in Education: Principles, Policy and Practice, 5(1), 111-122.

Herman, J. (1992). What research tells us about good assessment. Educational Assessment, 49(8), 74-78.

Hill, J. (1998). Rethinking assessment: The challenging issues for schools and teachers. Unpublished thesis presented in partial fulfilment of the requirements of the degree of Master of Educational Administration, Massey University, Albany, Auckland, New Zealand.

Hill, M. (1997a, 4-7 December). Models of assessment practices of teachers in New Zealand primary schools. Paper presented at the Annual Conference of the New Zealand Association for Research in Education, Auckland University, Auckland, New Zealand.

Hill, M. (1997b). Planning valid and valuable school assessment systems: A possible model for New Zealand primary schools. Teachers and Curriculum, 1, 85-98.

Hill, M. (1998a). Recording and reporting action research (An action research project carried out for the Ministry of Education Final). Hamilton, New Zealand: School of Education, University of Waikato.

Hill, M. (1998b, 3-6 December). Using assessment to differentiate: A case study of the use of assessment specifically to improve learning. Paper presented at

the Annual Conference of the New Zealand Association for Research in Education, University of Otago, Dunedin, New Zealand.

Hill, M. (1999, 14-15 October). Dot, slash, cross: How school-based assessment for accountability can drive teachers to ticking instead of teaching. Paper presented at the Bay of Plenty, Waikato and Poverty Bay Intermediate Deputy and Assistant Principals' Conference, Mt Maunganui, New Zealand.

Hill, M. (2000). Dot, slash, cross: How assessment can drive teachers to ticking instead of teaching. SET(1), 21-25.

Hill, M., & Jones, M. (1995, 5-8 July). Partnership in teacher education. Paper presented at the Annual Conference of the Australian Teacher Education Association, Australian Catholic University, Sydney, Australia.

Holdaway, D. (1972). Independence in reading: A handbook on individualized procedures. Auckland, New Zealand: Ashton Educational.

Holt, J. (1974). How children fail. London: Penguin.

House, E. R. (1998). Schools for sale: Why free market policies won't improve America's schools, and what will. New York: Teachers' College Press.

Hultqvist, K. (1998). A history of the present on children's welfare in Sweden: From Froebel to present-day decentralization projects. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (pp. 91-116). New York: Teachers' College Press.

Johannesson, I. A. (1998). Genealogy and progressive politics: Reflections on the notion of usefulness. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (pp. 297-315). New York: Teachers' College Press.

Johnson, M. (1987). The body in the mind: The bodily basis of meaning, imagination and reason. Chicago: Chicago University.

Johnston, P. (1989). Constructive evaluation and the improvement of teaching and learning. Teachers' College Record, 90(Summer), 535-549.

Jones, A. (1992). Writing feminist educational research: Am I in the text? In S. Middleton & A. Jones (Eds.), Women and education in Aotearoa New Zealand 2 (pp. 18-32). Wellington, New Zealand: Bridget Williams Books.

Jupp, V., & Norris, C. (1993). Traditions in documentary analysis. In M. Hammersley (Ed.), Social research: Philosophy, politics and practice (pp. 37-51). London: Sage Publications.

Kamii, C. (1982). Evaluation: It all depends on your theory. Australian Journal of Early Childhood, 10(2), 3-9.

Kelsey, J. (1997). The New Zealand experiment: A world model for structural adjustment? (2nd ed.). Auckland, New Zealand: Auckland University Press with Bridget Williams Books.

Kemmis, S., & McTaggart, R. (1988). The action research planner. Geelong, Victoria, Australia: Deakin University Press.

Kirk, J., & Gilmore, A. (1995). Essential skills: Identification, teaching, assessment and reporting. Christchurch, New Zealand: Canterbury University.

Kohl, H. (1969). The open classroom: A practical guide to a new way of teaching. New York: Vintage Books.

Kvale, S. (1996). InterViews: An introduction to qualitative research interviewing. Thousand Oaks, CA: Sage Publications.

Lange, D. (1988). Tomorrow's schools: The reform of education administration in New Zealand. Wellington: Government Printer.

Lather, P. (1986). Research as praxis. Harvard Educational Review, 56(3), 257-277.

Lather, P. (1991). Getting smart: Feminist research and pedagogy with/in the postmodern. New York: Routledge.

Lauder, H. (1990). The New Right revolution. In S. Middleton, J. Codd, & A. Jones (Eds.), New Zealand educational policy today: Critical perspectives (pp. 1-26). Wellington, New Zealand: Allen and Unwin.

Lauder, H., Middleton, S., Boston, J., & Wylie, C. (1988). The third wave: A critique of the New Zealand Treasury's report on education. New Zealand Journal of Education Studies, 23(1), 15-33.

Lee, G., & Lee, H. (1992). Examinations and the New Zealand school curriculum past and present. Delta Research Monograph, 12, 1-53.

Lee, H. (1987). The junior civil service examination reconsidered: A study of the changing function of a competitive examination: 1900-1912. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 107-121). Wellington, New Zealand: New Zealand Council for Educational Research.

Lee, H., & Lee, G. (1998, 3-7 December). Raising 'national standards': Some inconsistencies in the green paper on assessment for success in primary schools. Paper presented at the Annual Conference of the New Zealand Association for Educational Research, Dunedin, New Zealand.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage Publications.

Lingard, B., Knight, J., & Porter, P. (Eds.). (1993). Schooling reform in hard times. London: The Falmer Press.

Linn, R. (1993, April). Criterion-referenced measurement: A celebrated and frequently elaborated concept needed for the 21st century. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, USA.

Linn, R. (1994). Performance assessment: Policy, promises and technical measurement standards. Educational Research, 23(9), 4-14.

Livingstone, I. D. (1994). The workloads of primary school teachers. Wellington, New Zealand: Chartwell Consultants and NZEI.

Luke, A. (1992). The body literate: Discourse and inscription in early literacy training . (Unpublished paper)

Lyotard, J.-F. (1984). The postmodern condition: A report on knowledge (G. Bennington and B. Masumi, Trans.). Minneapolis: University of Minnesota Press.

MacDonald, F. (1997). Poor performance. Metro, 88(February), 58-65.

Madaus, G. F. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education, Part 1 (pp. 83-121). Chicago: University of Chicago Press.



Martin, R. (1988). Truth, power, self: An interview with Michel Foucault. In L. Martin, H. Gutman, & P. Hutton (Eds.), Technologies of the self: A seminar with Michel Foucault (pp. 9-15). London: Tavistock Publications.

Masters, G., & Forster, M. (1996). Progress maps. Melbourne: Australian Council for Educational Research.

Matthews, K. (1987). The long and winding road: The role of the nineteenth century school inspector: A case study. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 147-158). Wellington, New Zealand: New Zealand Council for Educational Research.

McCulloch, G. (1987). History and the curriculum. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 8-20). Wellington, New Zealand: New Zealand Council for Educational Research.

McHoul, A., & Grace, W. (1998). A Foucault primer: Discourse, power and the subject (2nd ed.). Dunedin: University of Otago Press.

McInerney, D. M., & McInerney, V. (1998). Educational psychology: Constructing learning (2nd ed.). Sydney: Prentice Hall.

McKenzie, D. (1984). Ideology and history of education in New Zealand. New Zealand Journal of Educational Studies, 19(1), 2-9.

McKenzie, D. (1987). The growth of school credentialling in New Zealand, 1878-1900. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 82-106). Wellington, New Zealand: New Zealand Council for Educational Research.

McKenzie, D. (1988). Little and lightly: The New Zealand Department of Education and mental testing 1920-1930. In M. Olssen (Ed.), Mental testing in New Zealand: Critical and oppositional perspectives (pp. 76-89). Dunedin, New Zealand: University of Otago Press.

McKenzie, N. R. (1930). School surveys 1927 and 1929. New Plymouth, New Zealand: North and South Taranaki Branches of the New Zealand Educational Institute.

McNeil, L. M. (1986). Contradictions of control: School structure and school knowledge. New York: Routledge and Kegan Paul.

McNiff, J. (1988). Action research: Principles and practice. London: Routledge.

Meadmore, D. (1993). For reasons of governmentality: A genealogy of dividing practices in Queensland schooling. Unpublished Doctor of Philosophy, University of Queensland, Brisbane.

Meadmore, D. (1995). Linking goals of governmentality with policies of assessment. Assessment in Education: Principles, Policy and Practice, 2(1), 9-22.

Meadmore, D., Limerick, B., Thomas, P., & Lucas, H. (1995). Developing practices: Managing the managers. Journal of Education Policy, 10(4), 399-411.

Melser, J. (1981). Who's going to lick the bowl? Auckland, New Zealand: Shortland Publications.

Merriam, S. B. (1988). Case study research in education: A qualitative approach. San Francisco: Jossey-Bass.

Metge, J., & Kinloch, P. (1978). Talking past each other: Problems of cross cultural communication. Wellington, New Zealand: Victoria University Press/Price Milburn.

Middleton, S. (1987). Feminism and education in post-war New Zealand: An oral history. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 132-146). Wellington, New Zealand: New Zealand Council for Educational Research.

Middleton, S. (1988). Researching feminist educational life histories. In S. Middleton (Ed.), Women and education in Aotearoa (pp. 68-93). Wellington, New Zealand: Allen and Unwin.

Middleton, S. (1990). Women, equality, and equity in liberal educational policies 1945-1988: A feminist critique. In S. Middleton, J. Codd, & A. Jones (Eds.), New Zealand education policy today (pp. 68-93). Wellington, New Zealand: Allen and Unwin.

Middleton, S. (1998). Disciplining sexuality: Foucault, life histories and education (1st ed.). New York: Teachers' College Press.

Middleton, S. (1999, 22-23 February). Sitting in rows and teaching on-line: Life-histories, technology and pedagogy. Paper presented at the Human Sciences Research Council Workshop, Strategies for Building Research Capacity among Women at Tertiary Institutions in South Africa, Pretoria, South Africa.

Middleton, S., & May, H. (1997). Teachers talk teaching 1915-1995: Early childhood, schools and teachers' colleges. Palmerston North, New Zealand: Dunmore Press.

Miller, P., & Rose, N. (1990). Governing economic life. Economy and Society, 19(1), 1-31.

Minister of Education. (1878). First annual report of the Minister of Education. AJHR, H-1, 1-116.

Ministerial Working Party on Assessment for Better Learning. (1990). Tomorrow's standards. Wellington, New Zealand: Learning Media.

Ministry of Education. (1991). The achievement initiative. The New Zealand Education Gazette, 70(7), 1-2.

Ministry of Education. (1992). National workshop on professional development in educational assessment. Stella Marris Conference Centre, Wellington, New Zealand: Ministry of Education.

Ministry of Education. (1993a). The New Zealand curriculum framework. Wellington, New Zealand: Learning Media, Ministry of Education.

Ministry of Education. (1993b). School charters and the revised national education guidelines. New Zealand Education Gazette, 72(7), 3-4.

Ministry of Education. (1994). Assessment: Policy to practice. Wellington, New Zealand: Learning Media.

Ministry of Education. (1998). Assessment for success in primary schools. Wellington, New Zealand: Ministry of Education.

Ministry of Education. (1999a). Information for better learning: National assessment in primary schools: Policies and proposals (final draft). Wellington, New Zealand: Ministry of Education.

Ministry of Education. (1999b). Revised national education guidelines. New Zealand Education Gazette, 78(22), 13.

Morgan, D. (1981). Men, masculinity and the process of sociological enquiry. In H. Roberts (Ed.), Doing feminist research (pp. 83-113). London: Routledge and Kegan Paul.

Moss, L. (1989). American influences on New Zealand education 1840-1945. In D. Philips, G. Lealand, & G. McDonald (Eds.), The impact of American ideas on New Zealand's educational policy, practice and thinking (pp. 34-49). Wellington, New Zealand: NZ-US Educational Foundation and New Zealand Council for Educational Research.

Munro, P. (1998). Subject to fiction: Women teachers' life history narratives and the cultural politics of resistance. Buckingham, U.K.: Open University Press.

Murphy, R., & Torrance, H. (1988). The changing face of educational assessment. Milton Keynes, UK: Open University Press.

National Education. (1919). The official journal of the New Zealand Educational Institute. National Education, 1(1), 1.

National Education. (1920). Intelligence tests. National Education, 2(21), 403.

New Zealand Treasury. (1984). Economic management: Brief to the incoming government. Wellington, New Zealand: Government Printer.

New Zealand Treasury. (1987). Government management: Brief to the incoming government 1987 (Vol. 2). Wellington, New Zealand: The Treasury.

Nias, J. (1989). Primary teachers talking. London: Routledge and Kegan Paul.

Nisbet, J. (1992). Assessment and curriculum reform. Aberdeen: University of Aberdeen.

Nisbet, J. (1993, 5-8 April). Educational assessment and curriculum reform. Paper presented at the OECD, Paris.

Oakley, A. (1981). Interviewing women: A contradiction in terms. In H. Roberts (Ed.), Doing feminist research (pp. 30-61). London: Routledge and Kegan Paul.

Olssen, M. (1987). What really happened? Varieties of educational history. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 21-38). Wellington, New Zealand: New Zealand Council for Educational Research.

Olssen, M. (1988). Introduction: Mental testing in New Zealand. In M. Olssen (Ed.), Mental testing in New Zealand: Critical and oppositional perspectives (pp. 9-25). Dunedin, New Zealand: University of Otago Press.

Olssen, M., & Morris Matthews, K. (1995). Education, democracy and reform: An introduction. In M. Olssen & K. Morris Matthews (Eds.), Education, democracy and reform (pp. 1-16). Auckland, New Zealand: The University of Auckland.

Olssen, M., & Morris Matthews, K. (Eds.). (1997). Education policy in New Zealand: The 1990s and beyond. Palmerston North, New Zealand: The Dunmore Press.

Openshaw, R. (1987). Introduction. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 1-7). Wellington, New Zealand: New Zealand Council for Educational Research.

Paris, S. G., & van Kraayenoord, C. E. (1992). New directions in assessing children's reading. Psychological Test Bulletin, 5(1), 20-27.

Patton, M. Q. (1990). Qualitative evaluation and research methods (2nd ed.). London: Sage Publications.

Perkins, D. (1995). Outsmarting IQ: The emerging science of learnable intelligence. New York: The Free Press.

Peshkin, A. (1988). Virtuous subjectivity: In the participant observer's I's. In D. N. Berg & K. K. Smith (Eds.), The self in social enquiry: Researching methods (pp. 267-268). Newbury Park, CA: Sage Publications.

Peters, M. (1992). Reform and the politics of the national curriculum. Unpublished paper. University of Canterbury.

Peters, M. (1995). Education and the postmodern condition: Revisiting Jean Francois Lyotard. Journal of Philosophy in Education, 29(3), 387-400.

Peters, M., Marshall, J., & Massey, L. (1994). Recent educational reforms in Aotearoa. In E. Coxon, K. Jenkins, J. Marshall, & L. Massey (Eds.), The politics of learning and teaching in Aotearoa-New Zealand (pp. 251-272). Palmerston North, New Zealand: The Dunmore Press.

Phillips, D. (1998, 3-7 December). National assessment policy and the role of national tests. Paper presented at the Annual Conference of the New Zealand Association for Research in Education, Dunedin, New Zealand.

Pole, C. J. (1993). Assessing and recording achievement. Buckingham, UK: Open University Press.

Polkinghorne, D. (1983). Methodology for the human sciences: Systems of inquiry. Albany, NY: State University of New York Press.

Pollard, A., Broadfoot, P., Croll, P., Osborn, M., & Abbott, D. (1994). Changing English primary schools? The impact of the Education Reform Act at Key Stage 1. London: Cassell.

Popkewitz, T. S., & Brennan, M. (1998). Restructuring of social and political theory in education: Foucault and a social epistemology of school practices. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (1st ed., pp. 3-35). New York: Teachers' College Press.

Postman, N., & Weingartner, C. (1969). Teaching as a subversive activity. New York: Celacorte Press.

Poulson, L. (1996). Accountability: Keyword in education reform discourse. Journal of Education Policy, 11(5), 579-592.

Preece, P., & Skinner, N. (1999). The national assessment in science at Key Stage 3 in England and Wales and its impact on teaching and learning. Assessment in Education: Principles, Policy and Practice, 6(1), 11-26.

Price, H. (1987). Reading books and reading in New Zealand schools 1877-1900. In R. Openshaw & D. McKenzie (Eds.), Reinterpreting the educational past: Essays in the history of New Zealand education (pp. 181-192). Wellington, New Zealand: New Zealand Council for Educational Research.

Pryor, J., & Torrance, H. (1996). Teacher-pupil interaction in formative assessment: Assessing the work or protecting the child? The Curriculum Journal, 7(2), 205-226.

Qualitative Solutions and Research Pty Ltd. (1997). QSR NUD\*IST 4 user guide (Version 2nd). Thousand Oaks, CA: Scolari, Sage Publications.

Radnor, H., & Shaw, K. (1995). A collaborative approach to moderation. In H. Torrance (Ed.), Evaluating authentic assessment (pp. 124-143). Buckingham: Open University Press.

Radnor, H. A. (1996). Assessment and control at Parkview School: a qualitative case study of implementing assessment change in a secondary school. Cresskill, NJ: Hampton Press.

Ramsay, P. (1987). The ethnographic approach to studying schools and classrooms. New Horizon, 28(November), 49-60.

Ramsay, P., Harold, B., Hill, D., Lang, C., & Yates, R. (1995). Final report of the school-based curriculum development project (Report commissioned by the Ministry of Education ). Hamilton, New Zealand: University of Waikato.

Reid, N., & Elley, W. B. (1991). Progressive achievement tests of reading: Teacher's manual. Wellington, New Zealand: New Zealand Council for Educational Research.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford & M. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 37-75). Boston: Kluwer Academic.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 445-462). Thousand Oaks, CA: Sage Publications.

Robertson, S., Dale, R., Thrupp, M., Vaughan, K., & Jacka, S. (1997). A review of ERO: Final report to the PPTA . Auckland, New Zealand: Auckland Uniservices Ltd.

Robertson, S. L. (1999). Strip away the bark. Expose the heartwood. Get to the heart of the matter: Re/regulating teachers' labour in New Zealand. New Zealand Journal of Educational Studies, 34(1), 121-132.

Rouse, A., & Dick, M. (1994). The use of NUD\*IST, a computerized analytical tool to support qualitative information systems research. Information Technology and People, 7(3), 50-62.

Sadler, R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18, 119-144.

Scheurich, J. J. (1997). Research method in the postmodern. London: The Falmer Press.

Schon, D. A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Schwab, J. (1969). College curriculum and student protest. Chicago: University of Chicago.

Scott, R. (1999). The red heart. Auckland, New Zealand: Vintage.

Sebatane, E. (1998). Assessment and classroom learning: A response to Black and Wiliam. Assessment in Education: Principles, Policy and Practice, 5(1), 123-130.

Selvaraj, S. D. (1996, 18 September). Using NUD\*IST for qualitative data analysis in accounting research. Paper presented at the Staff Seminar, Department of Accounting, University of Waikato, Hamilton, New Zealand.

Shepard, L. (1989). Why we need better assessments. Educational Leadership, 46(April), 4.

Shepard, L. (1991). Will national tests improve student learning? Phi Delta Kappan, 71, 232-238.

Sherman, R. R., & Webb, R. B. (1988). Research methods in education: Focus and method. London: The Falmer Press.

Shuker, R. (1987). The one best system? A revisionist history of state schooling in New Zealand. Palmerston North, New Zealand: The Dunmore Press.

Simola, H. (1997, 27-30 August). From exclusion to self-selection: Citizenship and examination of behaviour in Finnish elementary schooling from



the 1860s to the 1990s. Paper presented at the European Sociological Association Conference, 20th Century Europe: Inclusions/Exclusions, University of Essex, UK.

Simola, H., Heikkinen, S., & Silvonen, J. (1998). A catalog of possibilities: Foucauldian history of truth and educational research. In T. S. Popkewitz & M. Brennan (Eds.), Foucault's challenge: Discourse, knowledge and power in education (pp. 64-90). New York: Teachers' College Press.

Smart, B. (1985). Michel Foucault. Milton Keynes: The Open University.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. Educational Researcher, 15(1), 4-12.

Spradley, J. P. (1979). The ethnographic interview. New York: Holt, Rhinehart and Winston.

Stake, R. E. (1994). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 236-247). Thousand Oaks, CA: Sage Publications.

Stake, R. E. (1995). The art of case study research. London: Sage Publications.

Stenhouse, L. (1975). An introduction to curriculum research and development. London: Heinemann.

Stiggins, R. J. (1997). Student-centred classroom assessment (2nd ed.). Upper Saddle River, New Jersey: Prentice-Hall.

Stoll, L. (1998, 1-5 February). Supporting school improvement. Paper presented at the First Follow-up Conference on the OECD activity "Combating Failure at School", Christchurch, New Zealand.

Strachan, J. (1997). Feminist educational leadership in a "New Right" context in Aotearoa/New Zealand. Unpublished doctoral thesis, The University of Waikato, Hamilton, New Zealand.

Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, California: Sage Publications.

Sutton, R. (1991). Assessment: A framework for teachers. London: Routledge.

Sutton, R. (1994). Striking similarities. Education (28 October), 334-335.

Thackery, S., Syme, K., & Hendry, D. (1992). A survey of school entry practices: How schools gather information on new entrants. Wellington, New Zealand: Learning Media, Ministry of Education.

The Commission on Education. (1962). Report of the Commission on Education in New Zealand. Wellington: Government Printer.

Thrupp, M., Harold, B., Mansell, H., & Hawksworth, L. (2000). Mapping the cumulative impact of educational reform: A study of seven New Zealand schools. Hamilton, New Zealand: School of Education, University of Waikato.

Thrupp, M., & Smith, R. (1999). A decade of ERO. New Zealand Journal of Educational Studies, 34(1), 186-198.

Tierney, R. J., Crumpler, T. P., Bond, L., & Bertelsen, A. (1996, April 12). Aligning report cards with assessment reform efforts in literacy: Studies of shifting philosophies and practices. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Timperley, H. S., & Robinson, V. M. (1997). The micropolitics of accountability: The case of staff appraisal. In R. Macpherson (Ed.), The politics of accountability: Educational and international perspectives (pp. 157-171). Thousand Oaks, CA: Corwin Press.

Torrance, H. (Ed.). (1995). Evaluating authentic assessment. Buckingham, U.K.: Open University Press.

Torrance, H. (1998). Review essay. Assessment in Education: Principles, Policy and Practice, 5(2), 277-279.

Torrance, H., & Pryor, J. (1995). Investigating teacher assessment in infant classrooms: Methodological problems and emerging issues. Assessment in Education: Principles, Policy and Practice, 2(3), 305-320.

Velde, M. (2000). Tangata whenua involvement, [Internet]. New Zealand Education Gazette. [http://www.edgazette.govt.nz/articles/show\\_articles.cgi?id=5745](http://www.edgazette.govt.nz/articles/show_articles.cgi?id=5745) [2000, 14 August].

Viechnicki, K., Barbour, N., Shaklee, B., Rohrer, J., & Ambrose, R. (1993). The impact of portfolio assessment on teacher classroom activities. Journal of Teacher Education, 44(5), 371-377.

Vygotsky, L. (1978). Mind in society. Cambridge, MA: Harvard University Press.

Vygotsky, L. (1986). Thought and language. Cambridge, MA: Harvard University Press.

Walkerdine, V. (1984). Developmental psychology and the child-centred pedagogy: The insertion of Piaget into early education. In J. Henriques, W. Hollway, C. Urwin, C. Venn, & W. Walkerdine (Eds.), Changing the subject: Psychology, social regulation and subjectivity (pp. 153-202). London: Methuen.

Walkerdine, V. (1992). Progressive pedagogy and political struggle. In C. Luke & J. Gore (Eds.), Feminisms and critical pedagogy (pp. 15-24). New York: Routledge.

Weedon, C. (1987). Feminist practice and poststructuralist theory. Cambridge, MA: Blackwell.

White, S. (1991). Political theory and postmodernism. Cambridge: Cambridge University Press.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership, 46(7), 41-47.

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.

Willis, D. (1992). Assessment trends in 1992: Future directions handicapped by a legacy from the past? In H. Manson (Ed.), Annual Review of Education in New Zealand (pp. 245-260). Wellington, New Zealand: Victoria University.

Willis, D. (1994). School-based assessment: Underlying ideologies and their implications for teachers and learners. New Zealand Journal of Educational Studies, 29(2), 161-174.

Wolcott, H. (1973). The man in the principal's office. New York: Holt, Rinehart and Winston, Inc.

Wolcott, H. F. (1994). Transforming qualitative data: Description, analysis, and interpretation. Thousand Oaks, Ca: Sage Publications.

Wolf, D. P., LeMahieu, P. G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. Educational Leadership, 49(8), 8-13.

Wood, R. (1986). The agenda for educational measurement. In D. Nuttall (Ed.), Assessing educational achievement. London: Falmer Press.

Woods, P., Jeffrey, B., Troman, G., & Boyle, M. (1997). Restructuring schools, reconstructing teachers: Responding to change in the primary schools. Buckingham: Open University Press.

Working Party on Improving Learning and Teaching. (1974). Educational Development Conference: Improving learning and teaching (Report of the Working Party on Improving Learning and Teaching). Wellington, New Zealand: Government Printer.

Wylie, C. (1994). Self-managing schools in New Zealand: The fifth year. Wellington, New Zealand: New Zealand Council for Educational Research.

Wylie, C. (1998, October). School self-management in New Zealand: How can it make a real difference? Paper presented at the NZCER Conference, Effective School Self-Management: What's Ahead?, Wellington, New Zealand.

Young, E. (1992). Seven Blind Mice. New York: Philomel Books.

Zuber-Skerritt, O. (1995). Models for action research. In S. Pinchen & R. Passfield (Eds.), Moving on: Creative applications of action learning and action research (pp. 2-29). Queensland, Australia: Action Learning, Action Research and Process Management.

# **APPENDIX A**

## **INFORMED CONSENT FORMATS AND LETTERS**

**The University of Waikato**  
**School of Education**  
**Department of Professional Studies**

### **INFORMATION SHEET**

#### **Developing Assessment Systems and Practices in Primary Schools**

##### ***Who is the researcher?***

Mary Hill, a doctoral student at the University, is investigating how teachers in primary schools use assessment in their classrooms.

##### ***Where can she be contacted?***

Mary can be contacted at the University of Waikato in the Department of Professional Studies. The phone number is 07-8384604, or, at home, 07-8273454.

##### ***What is the study about?***

The study is designed to investigate how teachers within specific school environments develop workable assessment practices to meet the requirements of the National Education Goals for which the main purpose should be to improve students' learning and the quality of learning programmes.

I intend to work with the teachers in the school as they address this issue themselves, and to support them in identifying the current situation in the school, deciding on the aspects requiring change or modification, planning approaches to

change, collecting information about the improvements, analysing the results and reflecting again upon possible new directions.

### ***What will the participants have to do?***

The participants in this study will be the principal and teachers, and, perhaps, members of the Board of Trustees and some parents of children at the school. The study will involve the researcher in documenting what happens as the school undertakes the process of self-review in terms of the assessment practices within the school and the planning and implementation of any changes thought necessary by the participants as a result of this process. It will involve the participants in discussions about assessment and, probably, some teacher development initiatives. For some teachers, including the principal, the professional development leader and the development team members, it will involve interviews with the researcher. Some classroom observations will also be involved at times.

### ***How much time will be involved?***

The majority of the time taken for the study would be during the time the teachers spend on the school review process itself. This would include time spent on activities such as staff meetings, teacher development activities and school review and planning strategies. Additional to these activities, for some teachers, would be time to take part in interviews with the researcher (up to an hour once per term).

### ***What can the participants expect from the researcher?***

The intention is to engage in a win-win situation for both the researcher and the school and teachers within it. For her part, the researcher would gain an insider view of change and development in assessment across the curriculum. For the teachers, the intention is to provide support to undertake an action research project to develop their own assessment knowledge, practices and systems. In my capacity as a facilitator/researcher I will be able to provide knowledge and experience in the action research process itself, feedback about the results of the

assessment changes to the teachers and staff as a whole on a regular basis, and a source of information about assessment and related issues.

Participants can expect the interviews to be recorded, to receive copies of the transcripts of their interviews for review and correction, classroom observation by the researcher, and requests to use specific material from the data collected in the final report of the study. They can expect to have their privacy protected and personal details and research data to be kept confidential to the researcher and her supervisors. To achieve this, the researcher will use code names for both the school and the teachers involved in the final report. Any information about students at the school will likewise be confidential to the researcher and her supervisors and code names will be used where necessary.

***If you take part in this study, you have the right to:***

- \* Refuse to answer any particular question, and to withdraw from the study at any time,
- \* Ask any further questions about the study that occur to you during your participation,
- \* Provide information on the understanding that it is completely confidential to the researchers. All information is collected anonymously, and it will not be possible to identify you in any reports prepared from this study.
- \* Be given access to a summary of the findings from the study when it is concluded.

# **The University of Waikato**

## **Human Research Ethics Committee**

### **Developing Assessment Systems and Practices in Primary Schools**

#### **Consent Form**

I have read the Information Sheet for this project and have had the details of the study explained to me. My questions about the study have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I also understand I am free to withdraw from the study at any time, or to decline to answer any particular questions in the study. I agree to provide information to the researcher on the understanding that it is completely confidential.

I wish to participate in this study under the conditions set out on the Information Sheet.

Signed: \_\_\_\_\_

Name: \_\_\_\_\_

Date: \_\_\_\_\_



The Principal  
Waimaka School

22 May, 1996.

Dear Katy,

I am writing as a result of the meeting held last Friday 17 May to discuss Waimaka School's plans for development in the area of student assessment. As discussed, this letter is to describe in more detail my aims in undertaking an action research project in assessment at your school. I would appreciate it if you could share this letter with your teaching staff and the Board of Trustees as you see fit. I also attach an information sheet and a copy of a consent form that I will ask participants in the research to sign before the project commences. These measures are to conform to the requirements of the School of Education Research Ethics Committee and to ensure that there are no misunderstandings from the outset of this process.

As a candidate for a doctoral degree at the University of Waikato, I am intending to undertake a study into assessment practices of teachers in up to four primary schools. I hope Waimaka School will be one of these. My preference is to achieve this by being invited to assist each school in an action research project (self-review of student assessment) designed by the schools themselves to develop their own assessment systems and practices. In this way I believe such a project would be tailored to meeting both the needs of the schools and my own research aims. I am particularly keen to be involved with a school such as yours which is planning school and teacher development in assessment over the 1996 year and requires an outside facilitator such as myself to share the lead in this development.

I attach an information sheet that explains what I wish to investigate, how participants would be involved and a copy of a consent form that I would need the members of your staff to sign in order for me to proceed with this project. I would be very happy to discuss the details of the research further either by phone or in person. Please don't hesitate to contact me for this purpose.

Thank you for your request for my help in undertaking the revision of your assessment policies, systems and practices and your consideration of my proposed research.

Yours sincerely

Mary Hill

The Principal  
Arawhata School

19 February, 1996.

Dear Terry,

I am writing as a result of the meeting held last Friday 16 February to discuss Arawhata School's plans for development in the area of student assessment. As discussed this letter is to describe in more detail my aims in undertaking an action research project in assessment at your school. I would appreciate it if you could share this letter with your school development lead team, your teaching staff, and the Board of Trustees as you see fit. I also attach an information sheet and a copy of a consent form that I will ask participants in the research to sign before the project commences. These measures are to conform to the requirements of the School of Education Research Ethics Committee and to ensure that there are no misunderstandings from the outset of this process.

As a candidate for a doctoral degree at the University of Waikato, I am intending to undertake a study into assessment practices of teachers in two primary schools. I hope Arawhata School will be one of these. My preference is to achieve this by being invited to assist each school in an action research project designed by the schools themselves to develop their own assessment systems and practices. In this way I believe such a project would be tailored to meeting both the needs of the schools and my own research aims. I am particularly keen to be involved with a school such as yours which is planning school and teacher development in assessment over the 1996 year and requires an outside facilitator such as myself to share the lead in this development.

I attach an information sheet that explains what I wish to investigate and how participants would be involved and a copy of a consent form that I would need the members of your staff to sign in order for me to proceed with this project. I would be very happy to discuss the details of the research further either by phone or in person. Please don't hesitate to contact me for this purpose.

Thank you for your request for my help in undertaking the revision of your assessment policies, systems and practices and your consideration of my proposed research.

Yours sincerely

Mary Hill

Principal  
Kotiritiri School

3rd February, 1998.

Dear Luke,

Thank you for agreeing to continue to take part in my research project. As you know, I am a candidate for a doctoral degree at the University of Waikato and am undertaking a study into assessment practices of teachers in New Zealand primary schools. As a result of my previous interviews with yourself and other teachers I would now like to observe your practice in action. This will take place in the week 23-27 February.

I have included an information sheet about the research and a consent form. I would appreciate it if you and a representative of the Board of Trustees could sign the consent form and return it to me before I begin the research work at your school. This is to conform to the requirements of the School of Education Research Ethics Committee and to ensure that there are no misunderstandings from the outset of this process.

I'm really looking forward to spending the week in your classroom. As I said, I'll be there about 8.15 am on Monday 23rd February ready to begin. I would be very happy to discuss the details of the research further either by phone or in person. Please don't hesitate to contact me.

Yours sincerely

Mary Hill

**TO: Primary and Intermediate Schools**

**RE: Research for Doctor of Philosophy Degree**

**Draft Title: Developing Assessment Systems and Practices  
in Primary Schools**

### **INFORMATION SHEET**

#### ***Who is the researcher?***

Mary Hill, a doctoral student at the University of Waikato, is investigating how primary schools improve their assessment systems and how teachers use assessment in their classrooms (including how they record and report assessment information).

#### ***Where can she be contacted?***

Mary can be contacted at the University of Waikato in the Department of Professional Studies. The phone number is 07-8384604, or 07-8273454 (home), 07-8277906 (fax), and mhill@waikato.ac.nz (email).

#### ***What is the study about?***

The study is designed to investigate how teachers within specific school environments develop workable assessment practices to meet the requirements of the National Education Goals for which the main purpose should be to improve students' learning and the quality of learning programmes.

She has been working with teachers in two schools as they address this issue themselves, and has supported them as they carried out their school review and development in the assessment area.

In this next phase of the research the aim is to gather information from schools taking part in the national trial of recording and reporting materials. The purpose of this phase is to gain a picture of teacher assessment practices in primary schools across different New Zealand primary school contexts and to compare the findings with those from the two case study schools. To do this, she will use information gathered in the trial and interview principals and teachers in the participating schools.

### ***What will the participants have to do?***

The participants in this phase of the study will be the principal and teachers at each participating school. The study will involve the researcher in documenting what assessment practices are used and how these change. It will involve the participants in interviews with the researcher.

### ***How much time will be involved?***

The majority of the time taken would be to take part in interviews (of up to an hour) with the researcher.

### ***What can the participants expect from the researcher?***

The intention is to engage in a win-win situation for both the researcher and the school and teachers within it. For her part, the researcher would gain an insider view of change and development in assessment across the curriculum. For the teachers, the intention is to provide support with trialing the new materials.

Following discussion with the principal, the researcher will invite teachers to be interviewed about assessment and evaluation. She will provide each teacher who

agrees to be involved with an information sheet and ask him/her to sign a consent form.

Participants can expect the interviews to be recorded, to receive copies of the transcripts of their interviews for review and correction, and requests to use specific material from the data collected in the final report of the study. They can expect to have their privacy protected and personal details and research data to be kept confidential to the researcher and her supervisors. To achieve this, the researcher will use code names for both the school and the teachers involved in the final report. Any information about students at the school will likewise be confidential to the researcher and her supervisors and code names will be used if any data about children are used.

***If you take part in this study, teachers have the right to:***

- \* Refuse to answer any particular question, and to withdraw from the study at any time,
- \* Ask any further questions about the study that occur to them during their participation,
- \* Provide information on the understanding that it is completely confidential to the researcher and her supervisors. All information is collected anonymously, and it will not be possible to identify you in any reports prepared from this study.
- \* Be given access to a summary of the findings from the study when it is concluded.

**If you have any further questions or concerns, please contact  
Mary for clarification.**

**The University of Waikato**  
**Human Research Ethics Committee**

**Developing Assessment Systems and Practices in Primary Schools**

**Consent Form**

We have read the Information Sheet for this project and, where requested, have had the details of the study explained to us. Our questions about the study have been answered to our satisfaction, and we understand that we may ask further questions at any time.

We also understand we are free to withdraw from the study at any time, or to decline to answer any particular questions in the study. We agree to provide information to the researcher on the understanding that it is completely confidential.

We wish to participate in this study under the conditions set out on the Information Sheet.

Signed: \_\_\_\_\_  
(Principal)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Signed: \_\_\_\_\_  
(For the Board of Trustees)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

P.O. BOX 409  
CAMBRIDGE

11 FEBRUARY 1999

Dear

At long last I have returned to the serious business of analysing the interviews I carried out so long ago. I have enclosed a copy of your interview transcript and would be most grateful if you could read it over and check that you are still happy that it represents your ideas and practices at that time.

I realise, however, that you may have changed your ideas or assessment since the interview took place. Please feel free to comment on any of these changes on a separate sheet.

I am lucky enough to have been granted special study leave to assist me to get this work finished. I would appreciate it if you could return this script with any changes or extra comments in the enclosed return addressed envelope as soon as possible. I will then be able to complete my analysis and move to writing up the findings. If you are interested in seeing a copy of the thesis when it is complete, please let me know and send me a convenient address for this purpose. It could be the end of the year, at least, before I reach this stage, however.

Thank you again for all your work and help with this thesis. I do appreciate your participation.

With sincere thanks,

Mary Hill



## **APPENDIX B**

### **INTERVIEW 1 QUESTIONS**

1. What do you think is “good” assessment?
2. What instructions do you get about what to assess, how much and how often?
3. Which assessment activities do you find of most value? Why?
4. What assessment activities do you find of least value? Why?
5. Have you had any input into the assessment policy? If so, what? If not, have you seen the policy?
6. What assessment that you are doing are you really pleased with? Why?
7. Are there any aspects of assessment that you feel quite uncomfortable about? If so, what and why?
8. Assuming you became a teacher in the “ideal” school, what assessment practices would you implement?
9. Can you give some examples of how you use assessment to improve children’s learning and learning programmes?
10. What assessment activities do you use to gather evidence for reporting student achievement to others?
11. Who decides what assessment is appropriate and necessary in your classroom?

## **APPENDIX C**

### **INTERVIEW 2 QUESTIONS**

1. In the first interview many of the teachers I spoke to told me that knowing where the children were at helped them to assist children to learn better. Would you agree? If so, how does this help you to improve learning?
2. Can you explain to me how you think learning takes place, and how assessment is part of this process?
3. Can you tell me how you learnt about the assessment practices you use? Where do you think your methods and your ideas about assessment come from?
4. How do you think the work your school has been doing on assessment so far this year has added to your knowledge about assessment or your confidence about it?
5. What's been the most valuable part of the staff development process so far?
6. Some teachers I interviewed indicated to me that they believed the assessments they were expected to carry out filtered down from above, often originating with the ERO. Is this your understanding of how assessment expectations come about?
7. What about the teacher appraisal system in this school? Do you believe there are links between teacher appraisal and your assessment practices and systems?

# **APPENDIX D**

## **INTERVIEW 3 QUESTIONS**

### **Case Study Teacher Interview Questions - Interview 3 - base questions**

#### **1. Ways of doing teacher assessment:**

Can you please describe for me how you carry out assessment in your classroom?

For example, how do you do it in English? Mathematics? Science?

#### **2. Ways of gathering evidence:**

How do you gather evidence about what children can do?

#### **3. Ways of making decisions about children:**

How do you know what to do next once you know what a child can or cannot do?

Can you tell me how you group children, if you do?

How do you monitor progress against the national achievement objectives?

#### **4. Ways of recording information:**

Can you tell/show me how you record the assessment information?

How do you record 'levels' of achievement?

Why do you record information in this way?

## **Modified questions for each teacher**

### **Marvin**

#### **1. Ways of doing teacher assessment:**

The last time we talked, you said that you were trying to improve the way you carry out your assessment. Can you please describe for me how you go about assessing in your classroom now, say in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

#### **2. Ways of gathering evidence:**

You also mentioned that you have to have 'evidence'.

Do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? what do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

#### **3. Ways of making decisions about children:**

Last time you described “just noting things down as you go and then using that for your planning, even on a day-to-day basis”. Can you tell me more about how you use assessment to make decisions about what and how to teach?

For example, how do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

#### **4. Ways of recording information:**

Can you tell me about/show me how you record your assessment information?

Do you record “levels” of achievement?

If so, how?

Why do you record information this way?

## **Rose**

### **1. Ways of doing teacher assessment:**

The last time we talked, you said that assessment helped you find a place to start with teaching and make decisions about different activities you might use with different children. Can you please describe for me how you go about assessing in your classroom to get the sort of information you need, say in written language or reading?

(Prompt for other areas, like maths, social studies, science; and to find out more about these strategies, eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

### **2. Ways of gathering evidence:**

When we last talked you mentioned that you don't always write everything down, but you may have some ways that you collect evidence of learning.

Do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? what do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

### **3. Ways of making decisions about children:**

I was interested that you used assessment to plan some of your teaching. Can you tell me more about how you use assessment to make decisions about what and how to teach? (For example, how do you know what the next teaching step might be?)

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

### **4. Ways of recording information**

Can you tell me about/show me how you record your assessment information?

Do you record "levels" of achievement?

If so, how?

Why do you record information this way?

**Lisa**

**1. Ways of doing teacher assessment:**

The last time we talked, you said that the quicker you could find out where children are at, the more beneficial it is to both the child and yourself. Can you please describe for me how you go about assessing in your classroom now, say in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

**2. Ways of gathering evidence:**

You also mentioned that you like to know your children well and that you take notice of their interests and abilities during the year.

Do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? What do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

**3. Ways of making decisions about children:**

Last interview you said that it's important to find out what the children know so you know what to plan for in terms of teaching. Can you tell me more about how you use assessment to make decisions about what and how to teach?

For example, how do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

**4. Ways of recording information:**

Can you tell me about/show me how you record your assessment information?

Do you record "levels" of achievement?

**If so, how?**

**Why do you record information this way?**

## **Anne**

### **1. Ways of doing teacher assessment:**

The last time we talked, you said that you were assessing constantly and use an assessment book. Can you please describe for me exactly how you go about assessing in your classroom now, say in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

### **2. Ways of gathering evidence:**

Do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? what do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

### **3. Ways of making decisions about children:**

Last time you described briefly about your assessment book and the way you use it for planning. Can you tell me more about how you use assessment to make decisions about what and how to teach?

For example, how do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

### **4. Ways of recording information**

Can you tell me more about how you record your assessment information?

Do you record “levels” of achievement?

If so, how?

Why do you record information this way?



## **Rosemary**

### **1. Ways of doing teacher assessment:**

The last time we talked, you said that when you know where children are at you can start from that point and just go for it rather than “twaddling around”. I’d like to know more details about how you find where they are at?

Can you please describe for me how you go about assessing in your classroom now, say in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, .eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

### **2. Ways of gathering evidence:**

You stated last time that children have to be able to use knowledge as well as just knowing the facts. I'd like to know how you assess this, for example, do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? what do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

### **3. Ways of making decisions about children:**

Once you’ve collected the evidence, can you tell me about how you make decisions about what and how to teach next?

For example, how do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

### **4. Ways of recording information:**

Can you tell me about/show me how you record your assessment information?

Do you record “levels” of achievement?

If so, how?

Why do you record information this way?

## **Chris**

### **1. Ways of doing teacher assessment:**

The last time we talked, you said that you often find out what children know while you're working with them and from marking their work. I'd like to know more about your assessment techniques.

Can you please describe for me how you go about assessing in your classroom now, say in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, eg. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

### **2. Ways of gathering evidence:**

As well as marking work, do you collect any evidence of learning during assessment?

(If yes) what sort of evidence? what do you do with this evidence?

(prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

### **3. Ways of making decisions about children:**

Once you've collected the evidence, can you tell me about how you make decisions about what and how to teach next?

For example, how do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

### **4. Ways of recording information:**

Can you tell me about/show me how you record your assessment information?

Do you record 'levels' of achievement?

If so, how? Why do you record information this way?

## **Katy**

### **1. Ways of doing teacher assessment:**

The last time we talked, you said that you used pre- and post-tests in maths but I'd like to know a little more about how you go about your day-to-day assessment in the classroom. Can you please describe for me how you find out what the individual needs are, say, for example, in written language or reading?

(Prompt for other areas, like maths, social studies, science.)

(Prompt to find out more about these strategies, e.g. do you plan to this in advance? do you do this all the time or just at certain times of the day set aside for it? etc.)

### **2. Ways of gathering evidence:**

In finding out about these needs, do you collect any evidence?

(If yes) what sort of evidence? what do you do with this evidence?

(Prompts such as: Do you keep samples? When/how often? How do you store the evidence? Do you write it down? How?)

### **3. Ways of making decisions about children:**

You told me last time that you make teaching decisions based on your observations of what the children can and cannot do. For example, you suggested you work mainly with groups to meet teaching needs. Can you expand on this a little more, for instance:

How do you know what the next teaching step might be?

Do you use the national achievement objectives for this?

How do you monitor progress against the national achievement objectives?

### **4. Ways of recording information:**

Can you tell me about/show me how you record your assessment information?

Do you record 'levels' of achievement?

If so, how? Why do you record information this way?

# APPENDIX E

## TEACHER QUOTATIONS PILOT

The teacher I piloted this research strategy with was asked to say if the following quotations were *not at all like me*; *quite like me*; *very like me*.

### Ways of doing assessment

A        “I assess regularly. I've got the objectives in my book and I tick if they've got it, and I circle if it's not fully established. I note it down because then I know what I have to work on with those children.”

B.        “I don't do a lot of jottings so most of it's probably in my head. If I've got to get up and go over and write it down, I lose the thread of what I'm doing.”

C.        “I do a pre test using the objectives I want to teach. If they haven't grasped them then that's where I start. If they have then I either concentrate on a different objective or extend them on from the original one. I retest at the end of the unit.”

### Ways of gathering evidence about learning

D.        “I try and collect something every week so that by the end of the year it's quite a full record.”

E.        “I collect samples once a term, for instance, handwriting samples, examples of writing, proofreading, spelling and a maths checkpoint.”

F.        “I keep children's work for their portfolios, usually a good copy but occasionally I keep a sample in my folder as well.”

### Ways of making decisions about children's learning

G. "As I've gone along I've just taken notice of what happens in children's learning. For example, I've noticed that beginning sounds come first, then children will start listening to ending sounds, and then the middle. It's a matter of observing and seeing what they're doing and knowing what comes next."

H. "I use the assessment information to group them into three groups."

I. "I would pick up something I recognise, say in their writing, that several could do with reinforcement in that particular skill and we would do that."

J. "I think by the expectation of what you think the child is capable of. I constantly nudge that child to do that little bit extra or the little bit more in depth."

### Ways of recording information

K. "I have a book with all the objectives in it and I mark them off for each child as they've achieved a particular objective."

L. "About once a week I record using my roll book. All the objectives are laid out and I can see at a glance that a child has accomplished that objective."

M. "I guess it's incidental. Just a mental note or sometimes I scribble it on a scrap of paper or my workplan. When I have time; and in my individual records about twice a year."

### Using the curriculum achievement objectives

N. "I use the curriculum achievement objectives a little bit but they are rather broad so I break them down to be a lot more specific. Then I assess against those objectives."

O. "The objectives in my roll book are from the curriculum. That's what we're teaching to so I copied them, cut them out and stuck them in."

P. "In any unit planning, I usually list level 3 and level 4 achievement objectives. From that...I go back and look at each one maybe twice a year and think, well, yes, he was working at level 3."

### Self-assessment Strategies

Q. "I've got cards that are in the children's books and written in children's language. For example, "I can use full stops". Each goal that a child is working on is written on the card so they remember what they're trying to work on. They are taking responsibility for their own learning."

R. "The portfolios go home once a term and they're supposed to cover all the curriculum areas to show what's been done throughout the year. Parents really enjoy them and the children take pride in them. It's quite neat to pick one up and have a look at progress from the beginning of the year to the end."

S. "I don't use a lot of student self-assessment because the children are more concerned about counting and comparing their stickers and the time the conferencing takes is unmanageable."

## **APPENDIX F**

# **EXECUTIVE SUMMARY OF THE RECORDING AND REPORTING ACTION RESEARCH PROJECT CARRIED OUT FOR THE MINISTRY OF EDUCATION IN 1997-1998**

This report outlines the findings of an action research project that trialed the use of two recording and reporting cards in primary and intermediate schools and explored recording and reporting issues in a limited way in secondary schools. Over 130 schools throughout New Zealand participated in the project. These schools represented all deciles, were located in rural, small town, suburban and urban situations, and included schools of every size and kind.

The purpose of the trial was to evaluate the suitability of the recording and reporting materials in terms of the following criteria:

- explicit reference to the essential learning areas of the New Zealand Curriculum and, where appropriate, the various strands of the national curriculum statements;
- ease of interpretation;
- focus upon summary statement;
- usefulness for transferring information from one teacher to another, or one school to another, or from the school to the parent/caregiver;
- manageability;
- relevance for parents and students;
- application up to the end of Year 10 (Ministry of Education, 1997).

To investigate these issues, and more specifically, the practicability of the trial cards, a co-ordinator/researcher was contracted. She was responsible for: suggestions about the format of the trial cards; establishing the data base of schools involved in the project; distributing the trial materials to the participating schools; investigating existing practices in recording and reporting; supporting the implementation of the trial cards through offering a help-line, meetings and school

visits; evaluating the materials through questionnaires and interviews; investigating challenges faced by secondary schools; and reporting to the Ministry of Education on the findings. A research assistant was employed to assist with data entry and some analysis of the baseline and evaluation questionnaires.

The action research was carried out between April 1997 and June 1998. Cards were distributed to all participating schools, including the secondary schools, by June 30. Orders for the cards, as well as feedback about their inappropriateness, were received and dealt with during July 1997. Support was provided to primary/intermediate schools implementing the cards, and discussions were held with secondary schools about appropriate monitoring systems between July and December 1997. A newsletter to keep participating schools informed about progress and to introduce the evaluation aspect of the project was distributed in February 1998. The evaluation questionnaires were distributed towards the end of February and returned by 27 March, 1998. Between March and June, interviews were conducted with teachers and principals, and secondary schools were canvassed for feedback about how they were tackling the issues identified during 1997. In June 1998 the evaluation questionnaires and the interview data were analysed and the final report written.

Results of this action research indicated that a cumulative record card, similar to the one trialed but with modifications, would be very useful for schools who wished to use it to monitor progress against the achievement objectives and communicate this in a summarised form to new schools and teachers. Participating schools did not use the trial report card to the same extent. Indications were that most schools preferred to report to parents in their own way, although many found the trial card helpful in improving their own reports and suggested that a modified version of the trial report card be available on disk for customising by schools when appropriate. There was no support for using the report card as a cumulative record by retaining a photocopy of the report in the school. This was because schools saw the report and the cumulative record card as having different purposes and as being designed for different audiences. Feedback indicated that guidelines for keeping cumulative records and reporting



to parents, similar to those in *Keeping School Records* (Department of Education, 1989), are important to ensure that learning and teaching progress coherently throughout schooling.

During the project, the concept of computerisation of cumulative records in primary and intermediate schools was explored. Feedback from participating schools indicated that electronic systems can be very helpful in collating, analysing and presenting school-wide information about achievement. In contrast, participating primary and intermediate schools reported that individual students' cumulative record cards, such as those trialed, were simpler to keep and more manageable than using computer equipment for this purpose.

Results from the secondary schools participating in the action research indicated that, due to the organisational structure of secondary schools, cumulative record cards such as those used in the trial to monitor achievement were not feasible. In schools where students have specialist teachers for every subject, electronic systems appear necessary to keep track of individual students' progress across the curriculum. Traditionally, the only way these schools have monitored individual students' progress across the curriculum has been through school reports collated once or twice a year. Secondary schools in the project expressed a keen desire to improve their reporting and monitoring systems but required significant support, both technical and professional; to achieve this goal.

Primary, intermediate and secondary schools in this project commented upon the difficulty of recording achievement in levels when there is no consistency in the way the achievement objectives at each level are assessed.

The recommendations from this action research were made separately for two main types of school:

- recommendations about recording and reporting for primary, intermediate and middle schools, or, in other words, schools teaching children mainly as classes taught by one teacher; and,
- recommendations about recording and reporting for secondary schools, including those intermediate, area and Forms 1 to 7 schools where specialised subject teaching underpins the school organisation.

## Recommendations

### Primary, Intermediate and Middle Schools

#### 1. Cumulative record of achievement information

Based on the feedback from this action research project, a cumulative record of achievement card in both English and Māori should be provided for schools who choose to use it. (Details about how the trial cumulative record card should be modified for this purpose are contained in Chapter Six of the report.)

An electronic version in English and Māori should be provided.

The official requirements for keeping school enrolment and achievement records need to be clarified and communicated to all schools.

Guidelines for keeping the cumulative record of achievement, similar to those in *Keeping School Records* (Department of Education, 1989), should be provided to assist schools and teachers.

#### 2. Reporting to parents/caregivers

A modified version of the report card be made available in electronic format so that it can be customised and printed off as needed. This would give support to those schools requiring it and allow them to modify the design to suit the individual needs of their school and community. (Full details about the modifications recommended to the trial report card can be found in Chapter Six of the report.)

A Māori version should also be available in this way.

Guidelines for reporting, similar to those in *Keeping School Records* (Department of Education, 1989), should be provided to assist schools and teachers.

### **3. Guidelines for keeping cumulative school records and reporting to parents**

A resource booklet for recording and reporting should be produced which includes guidelines for keeping cumulative records of achievement and guidelines for reporting to parents and caregivers. Ways of supporting recording and reporting with a slim portfolio of work samples should also be outlined in this booklet. Professional development and support should also be provided, where necessary, to assist schools to implement a recording and reporting system.

### **4. National exemplar materials**

More national exemplar material needs to be provided. Examples are necessary for assessing student achievement and to help teachers decide whether the judgements they are making about student achievement are consistent with national expectations.

### **5. Computerisation of cumulative records in primary schools**

A modified version of the cumulative record of achievement and the report card be made available in electronic format so that they can be customised as needed.

## **Secondary Schools**

### **1. Assistance to develop recording and reporting systems**

Secondary schools should receive assistance to develop effective cumulative records for individual students which monitor achievement against the

achievement objectives, appropriate to the needs and assessment purposes of each school.

Assistance should include expertise to assist schools with facilitating professional development in assessment, recording and reporting strategies which contribute to improved learning.

A resource booklet and/or kit similar to that recommended by Nightingale (1993) should be produced to support secondary schools to monitor individual cumulative achievement.

## **2. National exemplar materials**

National exemplar material needs to be provided. Examples are necessary for assessing student achievement and to help teachers decide whether the judgements they are making about student achievement are consistent with national expectations. Exemplar materials should be provided to meet the needs of secondary teachers as well as primary/intermediate teachers.

In conclusion, the action research reported here confirmed that schools, both primary and secondary, would appreciate further support with recording and reporting. In particular, primary and intermediate schools requested the provision of a cumulative record card and guidelines for recording and reporting. In secondary schools, the results indicated guidelines as well as professional development to improve assessment, individual student monitoring and reporting up to the end of Year 10 and technical assistance to support schools introduce electronic monitoring systems are required. In order to increase the national consistency of recording and reporting achievement schools and teachers require more national exemplar material to inform the judgements they make about levels within the New Zealand Curriculum. Improving assessment practices, as well as the systems for recording and reporting achievement, will assist schools to ensure

that teaching builds on previous learning and prepares students for future learning experiences.

## References

Department of Education (1989). Keeping School Records. Wellington: Government Printer.

Ministry of Education (1997). Education Gazette, 29 March.

Nightingale, D. (1993). Records of Achievement Project Evaluation Report: Report on the production of records of achievement in a selection of New Zealand secondary schools in 1992. Blenheim.